

10 | 630992 PTO-892

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

REVISED VERSION

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
20 March 2003 (20.03.2003)

PCT

(10) International Publication Number  
WO 03/023654 A2

(51) International Patent Classification<sup>7</sup>: G06F 17/30

(21) International Application Number: PCT/US02/28277

(22) International Filing Date:  
6 September 2002 (06.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/317,735 6 September 2001 (06.09.2001) US

(71) Applicant (for all designated States except US):  
POLYVISTA, INC [US/US]; 1222 Ridgeley Drive,  
Houston, TX 77055 (US).

(71) Applicants and

(72) Inventors: ANWAR, Mohammed, Shahbaz [US/US];  
1222 Ridgeley Drive, Houston, TX 77055 (US). DA-  
HALE, Venkatesh [IN/US]; 5801 Spring Valley #1008,  
Dallas, TX 75240 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): STOUDT, Thomas  
[US/US]; 8902 Donys Drive, Houston, TX 77040 (US).

(74) Agent: STROZIER, Robert, W.; Strozier, Robert, W,  
PLL.C., P.O. Box 429, Bellaire, TX 77402-0429 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,  
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,  
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),  
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,  
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,  
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, ML, MR, NE, SN, TD, TG).

Published:

— with declaration under Article 17(2)(a); without abstract;  
title not checked by the International Searching Authority

(48) Date of publication of this revised version:  
30 October 2003

(15) Information about Correction:  
see PCT Gazette No. 44/2003 of 30 October 2003, Sec-  
tion II

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: A STRUCTURED DATABASE FOR DATA MINING RESULTS AND METHODS FOR CAPTURING, RETAINING  
AND ENHANCING KNOWLEDGE DISCOVERY

(57) Abstract:

WO 03/023654 A2

Best Available Copy

# PATENT COOPERATION TREATY

## PCT

### DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

(PCT Article 17(2)(a), Rules 13ter.1(c) and Rule 39)


Applicant's or agent's file reference <b>95626/14PCT</b>	IMPORTANT DECLARATION	Date of mailing(day/month/year) <b>27/03/2003</b>
International application No. <b>PCT/US 02/ 28277</b>	International filing date(day/month/year) <b>06/09/2002</b>	(Earliest) Priority date(day/month/year) <b>06/09/2001</b>
International Patent Classification (IPC) or both national classification and IPC <b>G06F17/30</b>		
Applicant <b>POLYVISTA, INC.</b>		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below

1. ☐ The subject matter of the international application relates to:
  - a. ☐ scientific theories.
  - b. ☐ mathematical theories
  - c. ☐ plant varieties.
  - d. ☐ animal varieties.
  - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes.
  - f. ☐ schemes, rules or methods of doing business.
  - g. ☐ schemes, rules or methods of performing purely mental acts.
  - h. ☐ schemes, rules or methods of playing games.
  - i. ☐ methods for treatment of the human body by surgery or therapy.
  - j. ☐ methods for treatment of the animal body by surgery or therapy.
  - k. ☐ diagnostic methods practised on the human or animal body.
  - l. ☐ mere presentations of information.
  - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art.
2. ☒ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:
 

☐ the description
☒ the claims
☐ the drawings
3. ☐ The failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions prevents a meaningful search from being carried out:
 

☐ the written form has not been furnished or does not comply with the standard.  
☐ the computer readable form has not been furnished or does not comply with the standard.
4. Further comments:

Name and mailing address of the International Searching Authority  European Patent Office, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  <b>Marja Brouwers</b>
--	---

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 203

All present claims relate to an extremely large number of possible apparatus or methods. In fact, the claims contain so many options and variables that a lack of clarity (and/or conciseness) within the meaning of Article 6 PCT arises to such an extent as to render a meaningful search of the claims impossible. Consequently, no search report can be established for the present application.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.5), should the problems which led to the Article 17(2) declaration be overcome.

-2-

groups, knowledge management portals, chats, threads, *etc.* While such procedures have individual merit, some of the drawbacks that diminish the value of these procedures over time include: (1) Decision Intelligence remains outside of the information system in generally unstructured and non-context specific formats, making it difficult to re-share or comprehend the gained intelligence without reconstruction and resulting in problems associated with "information islands", "information cataloging" and "information glut"; (2) Information redundancies are not removed and information re-discovery is more difficult; (3) Closed loop knowledge discovery, knowledge integration and historical trending are not possible or are not facilitate so that changing behaviors in pattern and exceptions cannot be gleaned; (4) As the information content grows, information searching becomes increasingly more difficult due to poor search mechanics across multiple systems and the requirements of human subjectivity; and (5) information saturation and/or overload can be a significant problem to knowledge accumulation and management.

[0005] Thus, there is a need in the art for improved structures and methods for harvesting information and knowledge from data, storing the information and knowledge in structures that facilitate memory, reuse and efficient knowledge adding, enhancing, revising, updating, deleting, augmenting, *etc.*

### **SUMMARY OF THE INVENTION**

[0006] The present invention provides a database adapted to acquire, store, and retain results from human and/or machine data mining activities so that information or knowledge can be derived and/or created therefrom allowing historical analysis and tracking and/or human commentary on the derived and/or created information or knowledge allowing decision makers to produce action plans utilizing the information or knowledge.

[0007] The present invention also provides a database structured to store non-redundant results from human and/or machine data mining activities, where the database includes information sufficient to reconstruct each stored results, information on each query that lead to each stored result and a structure sufficient to facilitate efficient retrieval, manipulation, integration, and distribution of the stored results.

[0008] The present invention provides a method for storing results from human and/or machine derived data mining results including the step of receiving a result from a human and/or machine data mining activity, comparing the received result to previously stored

-3-

results, storing the received result if the received result is new or updating a previously stored result if the received result is an augmentation, addition, correction or update of the previously stored result and if the received result causes a change in the database notifying a user group of the change. The method can also include the step of adding objective and/or subjective commentary, analysis, and/or rating. The method can also include the step of reviewing the results to minimize redundant results and adding further objective and/or subjective commentary, analysis, and/or rating by a group of specialized users. The method can also include the step of integrating the results into summary, collective, abstracted or compound results. The method can also include the step of running data mining routines against the results to derive higher order data patterns, trends or exceptions. The method also includes the step of deriving wisdom from the repository to be used in actions.

[0009] The present invention provides an interactive method for acquiring knowledge from data mining results including the steps of receiving input either manually via direct human entry into a knowledge database or automatically from participating data mining routines, storing the results in a temporary database area, screening the result against existing result entries, marking the results as either new, related or redundant, authoring the results with objective and/or subjective commentary, analysis and/or rating, submitting the authored result to the database for permanent storage, reviewing the results by a special group of users to add expert subjective and/or objective commentary, analysis and/or ratings, forming higher level results from the stored results, storing the higher level results in the database, removing redundant results from the database, presenting relevant results to specific user groups, running data mining routines against the results and integrating the results into patterns and trends and repeating the above steps on a continuous or periodic basis.

[0010] The present invention provides a knowledge storage and management system including an input interface that allows the results from one or more data mining routine running on one or more data sources (*e.g.*, OLAP cubes or other database) or from one or more human data mining activities to be screened for inclusion in an knowledge acquisition database which stores the results of the data mining activities with sufficient context information to recreate the data mining process and the source data structure. The system also includes knowledge review and pursue routines that permit the entry of human objective and/or subjective commentary to add additional informational content to the member of the

-4-

database, which are data exceptions or patterns identified by the data mining routines or humans. The system also includes specially tailored data mining routines that run against the knowledge database to produce higher level knowledge and informational content. The system can also includes information relating to the use of the knowledge or information in the system.

[0011] The present invention provides a system for converting data into information, information into knowledge and knowledge into wisdom. Information is extracted specific data directed to a given problem or interest, *i.e.*, context specific data, which can be textual or visual. Knowledge is the abstracted patterns or trends derived from information. Finally, wisdom is the integration of knowledge with human experience and insight and/or machine automated knowledge refinement.

#### **DESCRIPTION OF THE DRAWINGS**

[0012] The invention can be better understood with reference to the following detailed description together with the appended illustrative drawings in which like elements are numbered the same:

##### **Data Repository Structure**

[0013] Figure 1 depicts a schematic of an Entity Relationship database schema or diagram of a knowledge database repository of this invention;

[0014] Figure 2 depicts a schematic of a Refresh process for the database of Figure 1;

[0015] Figure 3 depicts a schematic of a PAS Analysis Server Update process for the database of Figure 1;

[0016] Figure 4 depicts a schematic of a User Query and Review Process for the database of Figure 1;

##### **Knowledge Management Methodology**

[0017] Figure 5A conceptually illustrates a preferred iterative and collaborative methodology of knowledge creation and sharing of this invention through various overlapping information qualification interactions;

[0018] Figure 5B illustrates a preferred embodiment of the Information Integration of Figure 5A, where relevant information from the repository of Figure 1 is aggregated and/or integrated with the base data;

[0019] Figure 5C depicts a conceptual flowchart of a preferred knowledge creation method

-5-

of this invention;

[0020] Figure 5D depicts the relationship between the primary informational entities pertaining to the framework;

#### **Repository Schema**

[0021] Figure 6A depicts a preferred embodiment of a repository wizard, which generates and records necessary metadata related to a result found in any OLAP cube;

[0022] Figure 6B depicts an example of an OLAP cube upon which the wizard of Figure 6A operates;

[0023] Figure 7A depicts a preferred partial database schema of this invention that is automatically created using the wizard of Figure 6A, which translates base data dimensions, hierarchy, levels of dimensions and measures into appropriate tables and columns, such that the exceptions and patterns related to a result are easily and uniquely stored in the database repository of this invention;

[0024] Figure 7B depicts the OLAP cube dimensional metadata recorded by the wizard of Figure 6A into the Repository\_Map table of Figure 7A;

[0025] Figure 7C depicts the OLAP cube security metadata that is automatically recorded by the wizard of Figure 6A into the "PW\_Constraint\_Definition" table of Figure 7A;

[0026] Figure 7D depicts an alternate preferred partial schema of this invention replacing multiple dimension specific tables with a one dimensional table, where details related to multiple dimensions are recorded in multiple records sharing the same repository;

[0027] Figure 7E depicts another alternate preferred partial schema of this invention having a single large de-normalized table, which persists content and context information all in one table;

[0028] Figure 7F depicts the Nugget Facts table of Figure 7A, which holds the Analytical information content of the Information Nugget;

[0029] Figure 7G depicts the Issues table of Figure 7A, which records Subjective issues relating to Information Nuggets;

[0030] Figure 7H depicts the Batch Tasks table of Figure 7A, which records the details related to any Evolved processes that publish potential Information Nuggets to the Repository;

[0031] Figure 7I depicts the Summary table of Figure 7A, which is one of many ways in which the contents of repository can be summarized to derive secondary intelligence on

-6-

subjects of interest on a periodic basis and allows for pursuing changing informational and exceptional subjects overtime;

### **Repository Interactions**

[0032] Figure 8A shows integrated and easy means of interacting with the repository, the wizard automatically reads the context of the anomaly and retrieves any existing information that may further explain or provide insights related to the current exception or pattern, thus the Information is readily shared without the need for explicit searching etc.;

[0033] Figure 8B shows a simple interface that presents information related to an existing Information Nugget;

[0034] Figure 8C shows a confirmation window associated with the Information Nugget submission wizard of this invention;

[0035] Figure 8D shows a simple interface that records objective and subjective commentary provided by the user;

[0036] Figure 8E shows a simple interface that records the objective score related to the information value of an Information Nugget or an Information Nugget Review;

[0037] Figure 9A shows a sample Information Nugget Fact table of this invention;

[0038] Figure 9B shows a sample entries from some Dimensional Tables of this invention that record Information Nugget context information;

[0039] Figure 9C shows an overview of the Information Nugget context categorization that occurs at storage time;

[0040] Figure 10A shows a "Drag n Drop" query interface, which is a part of the Analytical client; the user simply drags the dimensional members of interest into the query box and click on "Search", necessary query;

[0041] Figure 10B shows the corresponding SQL query generated by the search submitted in Figure 10A;

[0042] Figure 10C shows a flow chart of the query process of Figures 10A&B;

[0043] Figure 11 shows a simple Interface, which records user informational preference profiles to minimize Information Glut;

[0044] Figures 12A-G show a custom web-based interface built around the Repository to enable easy and customized information delivery, the database of this invention's self-documented and open architecture allows for such custom information delivery and



-7-

integration needs;

[0045] Figures 13 shows a sample interface where the entire contents of the repository can be organized in a dimensional tree like format; and

[0046] Figure 14 shows a example of Information Integration, where a Basket Analysis data mining routine has been customized to mine the repository.

#### **DETAILED DESCRIPTION OF THE INVENTION**

[0047] The inventors have found that a knowledge database can be constructed with a unique database schema to facilitate storing, authoring, updating, augmenting, searching, retrieving, reviewing, disseminating, commenting on and analyzing results of human and/or machine data mining activities and synthesizing the results into composite results, abstracted results, patterns, trends and/or other higher level information and/or knowledge. The inventors have also found that a methodology can be constructed for efficiently storage, updating, augmenting, searching, and retrieval of the results of human and/or machine data mining activities. Moreover, the inventors have found that a methodology can be constructed to capture human rating of results of human and/or machine data mining activities so that current information and historical information about results of data mining activities can be easily and efficiently identified, update, augmented, abstracted and otherwise turned into past, present and predicted future behavioral knowledge about an area of interest – business, scientific, medical, financial, governmental, intelligence, environmental, global, *etc.*

[0048] The key differentiating features of the present invention over the prior art include: (1) Self-documented, self-cataloging and open RDBMS based data repository; (2) Automatic generation, ease of administration and ease of change management mechanics; (3) Designed to facilitate knowledge creation, sharing and integration through flexible and extensible collaborative and knowledge query mechanisms; (4) User interface focuses on capturing knowledge through least and most relevant human input, while query mechanics focus on presenting most relevant information with least criteria specification requirements; (5) Enhanced user activity tracking and profiling helps with suggestive intelligence “This may be of interest to you!” metaphor; (6) Makes corporate knowledge pervasive and actionable; (7) Provides for summarization and abstraction of knowledge into larger insights or allows the construction of business wisdom; and (8) Storing information on each result that is sufficient to recreate the context of the given result. For additional information that regarding OLAP

-8-

databases, data mining routines, historical data storage and usage and other features disclosed herein, the reader is referred to United States and PCT Patent Application Serial Nos.: 08/721,899 filed 9/27/96 now USPN: 5,767,854 issued 6/16/98; 09/713,674 filed 15 November 2000 (corresponding PCT/US00/ 31523); 09/541,457 filed 3/31/00; 09/811008 filed 3/16/01 (corresponding PCT/US01/08467); and PCT/US02/19541 filed 6/19/02; PCT/US02/21538 filed 7/9/02, incorporated herein by reference.

[0049] The present invention broadly relates to a database having a unique database schema designed to facilitate the storage and use of human and/or machine generated data mining results (data exceptions, data anomalies, meta exceptions, data oddities, data patterns and/or data trends)

#### **Background**

[0050] The following diagrams display the entity relationship diagrams and data flow diagrams. The entity relationship diagrams show the logical data entities and their association to each other. The data flow diagrams show the interaction between processes and data-stores. The processes are displayed as squares, cubes as cubes, and the data-stores as cylinders.

#### **Business Requirements**

[0051] The knowledge database and processes of this invention are well suited to satisfy business requirements including: (1) Self-Maintaining: The ability for the system to keep itself current with cube changes with minimal human support; (2) History Retention: The data should keep nuggets and issues history even when cubes or dimension members no longer exist in the database; and (3) Efficient Querying: The data should be stored in a manner that provides quick and efficient data retrieval for user queries.

#### **Data Repository Structure**

[0052] Referring now to Figure 1, a Entity Relationship diagram, generally 100, is shown, which described a preferred embodiment of a database schema for using the knowledge repository database of this invention. The diagram 100 includes the following database entities or tables: Departments or Grouping of Users 102, Databases or Collections of multidimensional cubes 104, Tasks or Scheduled processes 106, Users or collections of valid users 108, Roles or grouping of user types 110, Dimensions or Dimensions contained in the multidimensional cubes 112, Cubes or Cubes registered in a multidimensional database

engine (*e.g.*, MS Analysis Services) 114, Task Logs or Details of task executions 116, Issues or Business problems 118, Nuggets or Cube tuples with interesting values 120, Query Logs or Information about user cube retrieval 122, Role Memberships or Users that belong to a role 124, Dimension Constraints or Access rules applied to dimensions 126, Cube Dimensions or Dimensions contained cubes 128, Nugget Issues or Issues that apply to nuggets 130, Nugget Reviews or Collaborative explanations about nuggets 132, Nugget Members or Dimension members contained in nugget tuples 134, and Dimension Members or Members that belong to a given dimension 136.

[0053] These entities or database tables are split into three functional groups. The first group comprises tables of relational storage of the multidimensional cubes, where the tables are: the Cubes 114, the Databases 104, the Cube Dimensions 128, and the Dimension Members 136. These tables are maintained by a process that keeps the data consistent with the cube members. If a cube member is deleted, the refresh process marks the member as "invalid" so that historical nugget information is retained, and the client tools of this invention will not attempt to query the values from the cube that include the "invalid" member(s) or orphans. The user can ask for orphaned nuggets. Nuggets are orphaned generally when something in the source database got changed. A utility can allow the user to detect orphaned nuggets and provide a vehicle to fix the orphaned nuggets.

[0054] The second group pertains to the processing of scheduled algorithms or data mining routines. The second group includes the Tasks 106, the Task Logs 116, the Issues 118, the Nuggets 120, the Nugget Issues 130, and the Nugget Members 134. The Tasks 106 are manually entered. The scheduled process reads the task and automatically generates the data for the rest of the entities.

[0055] The last functional group is the security group. This group determines which values can be displayed to groups of people and includes the Users 108, Roles 110, and Role Memberships 124.

[0056] The diagrams displays the following relationships: (a) Departments 102 are composed of Users 108; (b) Databases 104 store data for Dimensions 112 and Cubes 114; Tasks 106 are the subject of Task Logs 116, which create Issues 118 and Nuggets 120; (c) Users 108 are the subject of Query Logs 122 and belong to Role Memberships 124; (d) Roles 110 have Role Memberships 124 and Dimension Constraints 126; (e) Dimensions 112 are the subjects of

-10-

Dimension Constraints 126, contain Dimension Members 136, and belong to Cube Dimensions 128; (f) Cubes 114 contain Cube Dimensions 128; (g) Issues 118 are explained by Nugget Issues 130; (h) Nuggets 120 belong to Nugget Issues 130, are the subject of Nugget Reviews 132, contain Nugget Members 134, and are listed in Query Logs 122; and (i) Dimension Members 136 define Nugget Members 134. The system also includes an archival routine (not shown), which allows older or less used nuggets to be archived to enhance database performance.

[0057] Referring now to Figure 2, a schematic diagram of a refresh process of this invention, generally 200, is shown to include a cube refresh process 202 that reads data from various Data Marts 204 and spins the data into Process Cubes 206. The process 200 also includes a Refresh 208 extracts data from the Process Cubes 210 and updates Cubes 212, Dimensions 214, and Dimension Members 216.

[0058] Referring now to Figure 3, a schematic diagram of a PAS Analysis Server Update process invention, generally 300, is shown to include a PAS routine 302 that executes Tasks 304 and writes execution details to Task Logs 306. The term PAS represent any data mining routine or program that includes data mining routines or algorithms. The PAS routine 302 also writes XML 308 and writes Nuggets or Nugget Members 310. An Issue Generator 312 determines nugget groupings and writes Issues/Nugget Issues 314, and write Knowledge Management (KM) 316.

[0059] Referring now to Figure 4, a schematic diagram of a User Query and Review Process of this invention, generally 400, is shown to include a Nugget Query module 402 that determines a user's access from a Role Memberships module 404. The Nugget Query module 402 displays issues from an Issues module 406, and displays nuggets or nugget members from a Nuggets and Nugget Members module 408 to the end users according to their profile stored in a User Profile module 410. The Nugget Query module 402 allows the creation and updating of issues in the Issues module 406, of reviews in a Nugget Review module 412, and of the profiles in the Users Profile module 410.

#### **Knowledge Management Methodology**

[0060] Referring now to Figure 5A, a conceptual iterative and collaborative method of knowledge creation and sharing of this invention generally 500, is illustrated, which is supported by the framework through various overlapped information qualification

-11-

interactions.

[0061] Evolved Discovery (Pre-Defined Batch Data Mining Processes) 502 and Ad-Hoc Discovery (Human detection of Exception or Pattern with or without the use of Data Mining) 504 are any processes that identify exceptions or patterns that are of potential interest. These findings may be based on the base cube data or based on the existing contents of Wisdom Repository.

[0062] "Information Authoring" interactions 506 entail the process of further researching the findings of potential interest to qualify it as an Information Nugget and supplementing the objective finding with subjective and objective commentary. Information Review interactions 508 entail the study of existing Information Nuggets in the Wisdom repository and providing further explanations, contradictions and extensions to these Information Nuggets, these reviews are recorded separately as Information Nugget Reviews. Information Pursue iterations 510 entail the study of existing Information Nuggets and/or Information Nugget Reviews and providing objective ratings to its information relevance or value.

[0063] Together, the Information Authoring 506, Information Review 508 and Information Pursue 510 constitute the Knowledge Creation process 512. As the Information Nuggets are qualified and embellished with subjective and/or objective information during the various interactions 506-510, they are permanently persisted (until meaningful) in the data repository of this invention and continue to be reviewed, pursued and made actionable through Information Integration. While the Information Nuggets represent intelligence at the most granular form, they can be grouped and abstracted or summarized to be part of a larger issue referred to herein as Information Issues through manual and/or automated means for deriving a larger picture or insights.

[0064] Information Integration 514 entails various means of making the information content of the data repository of this invention more useful (by abstracting, summarizing etc.) and available (through various information channels and delivery mechanisms) to users and operational systems. The entire process is then iterated in a participation and usage process 516 that allows the data stored in the repository to be continually or periodically evolved into a more detailed intelligence platform from which improved business decisions can be made.

[0065] As soon as a potential pattern or exception is detected or discovered by a Data Mining routine and/or other Analytical Routines (Evolved Discovery 502) and through Ad-hoc

-12-

processes or manual activity (Ad-Hoc Discovery 504), users can simply right click on the pattern or exception and retrieve existing information in the context (variables) of the exception (anomaly) or pattern and members (subjects) currently being investigated.

[0066] In absence of existing information or where the user has a different perspective or opinion than the stored or discovered information, new information nuggets can be authored simply by entering comments and/or an objective score to a pattern or exception. Unstructured related reference resources may also be stored with these nuggets into the Repository during the Information Authoring interaction 506.

[0067] Any explanations, contradictions and extensions to these Information Nuggets can also be added as reviews to existing information nuggets during the Information Review interaction 508, information relevance can be objectively provided by information consumers during the Information Pursue interaction 510. As these Information Nuggets are qualified and embellished with subjective and objective information, they are permanently persisted (until meaningful) in the data repository and continue to be reviewed, pursued and made actionable through Information Integration.

[0068] While the Information Nuggets presents intelligence at the most granular level; they can be grouped, abstracted and/or summarized into larger Information Issues, results and/or conclusion through manual and/or automated means. As users pursue the Information Nuggets, the users provide objective score and/or Subjective commentary about the Information Nuggets. The routines that facilitate review of the Information Nuggets also keep track of user search and usage activity which is recorded in order to enhance the relevance of Information Nuggets and Information Issues and to provide a mechanics for most interesting Information Nuggets and/or Issues to be presented first to any particular user or user group or to all users in general. The information and/or knowledge contained in the data repository can then be integrated efficiently into operational systems or other informational systems and/or delivered through a variety of channels within an organization to maximize knowledge exposure and utility.

[0069] Referring now to Figure 5B, a preferred embodiment of an Information Integration method of this invention, generally 520, is shown. The method shown in Figure 5B provides a framework, whereby relevant information from the repository is aggregated (where meaningful) and/or integrated with base or source data, from which exceptions and/or patterns

-13-

were originally identified, and presented to the user. The framework allows users to have quick access to the all information about the exceptions and/or patterns ("Good, Bad and Ugly") over time using a single interface. Additionally, the framework provides information nuggets and/or summaries to linked Data Mining and/or other analytics to improve query construction, to improve quality of results, to improve analysis performance, and/or to improve other aspects of data mining. Looking again at Figure 5B, the framework 520 includes an information nugget data repository or nugget database 522 in two way communication with data mining or evolved discovery 502 and human ad-hoc discovery 504. The two data mining processes 502 and 504 and the repository 522 are in two communication with multidimensional data sources 524. Additionally, the data mining process 502 and 504 are in two way communication with constructed multidimensional (virtual) data sources 526, which were constructed from data in one or more actual sources 524 by the data mining routines or by direct human construction.

[0070] When a user interacts and/or uses the system, the system transparently records the interactions or usages. This historical interaction data is used by the system to create specific user profiles, user group profiles, organizational profiles, content profiles, or similar profiles. These profiles are in turn used by the system to present to users, user groups, or organizational groups meaningful summaries, distilled knowledge, and/or derived intelligence related to subjects of interest to each user, user group or organization group. Thus, the system can present the most interesting or relevant Information Nuggets, summaries, distilled knowledge, and/or derived intelligence to the user at login. These profiles allow the system to be knowledgeable about subject of interest to specific users, user groups or organization groups. Moreover, the system provides enhanced, simple, user friendly information query interfaces and routines that are designed to present to the user the most relevant informational content efficiently and easily. The profile information can be used by system administrators to further fine tune the content of the source databases.

[0071] Referring now to Figure 5C, a flowchart, generally 530, is shown that depicts a preferred method for knowledge creation and management. Batch patterns and/or exceptions 532 identified by various batch routines (not shown) are recorded to a temporary staging area 534, which may simply mean that the exception or pattern is stored in the main Information Nuggets Fact table marked as temporary. Once the batch pattern or exception 532 is stored

-14-

in the area 534, the pattern or exception 532 is pre-screened in a pre-screening step 536. The pre-screening step 536 subjects the batch pattern or exception to various automated and/or manual pre-screening tasks to check for Redundancy, Information Value, or the like. After pre-screening, the batch pattern or exception 532 is forwarded to a conditional step 538. User Pattern or Exception 540 identified by a user are not subject to pre-screening step 536 or held in the temporary storage area 534, but are forwarded directly to the conditional step 538. Regardless of the way the Exception or Pattern 532 or 540 were identified, a common information qualification process is used to create an Information Nugget. As soon as a potential patterns or exception is detected or discovered, users can simply right click and retrieve and study any existing information in the context (variable) of the exception or pattern and members (subjects) currently being investigated.

[0072] The conditional step 538 checks the repository to determine if an information nugget that is identical to or similar to the identified pattern or exception 532 or 540. The conditional step 538 provides for four different conditions. First, if the pattern or exception 532 or 540 already exists or is similar to an existing pattern or exception and the user wants to add objective ratings and/or subjective commentary to the existing nugget, then control is transferred along a YES branch 542 to an information nugget review step 544 which produces an information review document 546 that is added to the repository.

[0073] Second, if the pattern or exception 532 or 540 already exists or is similar to an existing pattern or exception and the user wants to objective rating the importance of the existing nugget, then control is transferred along a YES branch 548 to an information nugget pursue step 550 which produces an information pursue document 552 that is added to the nugget face table.

[0074] Finally, if the pattern or exception 532 or 540 does not exists or is not similar to an existing pattern or exception or the user has a different perception, thought or conclusion on the nugget and wants to create a new nugget, then control is transferred along a NO branch 554 to an information nugget authoring step 556 which produces an information nugget 558 that is added to the repository. In the case of an existing information nugget, where the user has a different perspective, thought, conclusion or opinion from the existing nugget; a new information nuggets is authored by simply entering comments and objective score, unstructured related reference resources may also be persisted with these nuggets into the



Repository during the Information Authoring step or interaction.

[0075] Any explanations, contradictions and extensions to these Information Nuggets can also be added as reviews to existing information nuggets during the Information Review process; while information relevance can be objectively added to the information nuggets by information consumers during the Information Pursue process. It is important to note that Information Review and Information Pursue processes are independent processes and do not have to occur in a particular sequence, *i.e.*, a user can invoke the Information Pursue process or the Information Review process independently or in order.

[0076] Referring now to Figure 5D, a relationship scheme or map, generally 570, is shown that depicts the nature and dimensionality of the communication pathways between the primary informational processes comprising the knowledge management system and framework of this invention. The map 570 starts with a result 572 from a data mining routine (evolved discovery) or a user (ad hoc discovery). Each result 572 from a data mining routine includes OLAP cube cell information including an unique Dimensional Context and an Analytical Context, which can include the type of analysis, *e.g.*, Basket Analysis, Deviation Analysis, User Driven Cross-Tab findings or the like. Each result 572, which comprises this combined information, can give rise to one or more Information Nuggets 574 or Information Nugget refinements via a pathway 576 as described in the process of Figure 5C above. Thus, the pathway 576 between the results 572 and the nuggets 574 is 1 to n or 1 to many mapping, where the notation 0,1, 0,n, or 0.m notation represents the cardinality of the pathway at each node or box.

[0077] Each Information Nugget 574 may have one or more Information Review 578 via pathway 580 (a 1 to n mapping) and/or Information Pursues 582 via pathway 584 (a 1 to n mapping) associated with it. Further, one or more Information Nuggets 574 may be grouped or abstracted to one or more Information Issues 586 via pathway 588 (a n to m mapping) and/or Information Actions 590 via pathway 592 (a n to m mapping). Each Information Review 578 is based on a Single Information Nugget 574 and may have one or more Information Pursue 582 via pathway 594 (a 1 to n mapping) associated with it. Each Information Pursue 582 is based on a Single Information Nugget 574 or Information Review 578 via pathways 594 and 584 (1 to n mappings). One or more Information Actions 590 (which is any initiative or action taken based on the Information Nuggets or Issues) can be

-16-

based on one or more Information Issues 586 via pathway 596 (a n to m mapping) and/or Information Nuggets 574 via pathway 592 (a n to m mapping). One or more Information Issues 586 can be based on one or more Information Nuggets 574 via pathway 592 (a n to m mapping) and can be any custom way of grouping the granular information into larger insights etc. and can intern be the basis of new Information Nuggets in future iterations. While the repository persists the details of these informational entities and/or processes, user friendly interfaces and routines facilitate the creation of and interaction with these informational entities and/or processes. Moreover, Information Nuggets 574 and Information Issues 586 can be further abstracted via pathways 598.

### **Repository Schema**

[0078] Referring now to Figure 6A, an example of an OLAP cube, generally 600, is shown to include 11 dimensions 602. and 7 measures 604. The cube 600 constitutes an example of a base or source data on which the ad hoc or evolved discovery processes operate, which can give rise to results that are presented to the system for retention (persisted) in the repository. As is discussed later in this document, a wizard automatically creates the necessary repository structures for each result including relevant cube data during the presentation and authoring steps of the entry process of the system. Looking at Figure 6B, a wizard input window, generally 650, of a preferred repository wizard of this invention is shown. The wizard generates and records the necessary metadata related to any OLAP cube. The wizard input window 650 includes database repository identification fields 652, a use integrated security check box 654 and a test connection button 656. The wizard input window 650 also includes source OLAP database informational fields 658, a connect button 660, and an only update cube security constraints check box 662. The wizard input window 650 also includes a maintain repository button 664, a build/update repository button 666 and a close button 668.

[0079] Referring now to Figure 7A, an example of a preferred partial database schema, generally 700, of this invention is shown. The schema 700 is illustrated for a source database for the retail food industry. Although the schema is illustrated using a source data from the food industry, it should be recognized that any source data can be used as well. The partial schema 700 includes at least a set 702 of source data identification table including dimensional tables 704 (educational\_level), 706 (sales\_customers), and 708 (sales\_time), measures table 710 and product table 712 (sales\_product). The remaining tables 714 are

-17-

information and knowledge associated tables. The information and knowledge tables 714 include at least the following tables: a nugget facts table 716, a repository map table 718, a constraints table 720, a batch tasks table 722, an information issues table 724, an issues map table 726, a search table 728, a users table 730, a batch task log table 732, an information review table 734, an information pursue table 736, an actions table 738 and an action map table 740.

[0080] The schema 700 is automatically created when the result being entered by the wizard of Figure 6B is submitted to the repository. The OLAP dimensions, hierarchy, levels of dimensions and measures are translated into entries in appropriate database repository tables and columns, such that the exceptions and patterns and the subject to which the exception or pattern relate may be easily, efficiently and uniquely recorded. Each Information Nugget is identified by a unique ID stored in an ID column in each table, which serve as a common link between the repository tables. In the above schema example, a separate table is built for every dimension in the cube and each field in the dimensional table corresponds to a level of the dimension. Each dimensional table has an additional column PW\_Written\_To\_Level, which records the level ordinal of the dimensional ordinal involved in the Information Nugget. Some of the other tables are Non-Cube specific and are common to several cubes. The schema can be ported to any ODBC compliant RDBMS and is completely open and extensible.

[0081] When a result that includes a dimension that has members such as a dimension with hierarchical children or a user submits a query with a keyword such as "Drink" that has children, then the system can also look for all nuggets that contain not only the high level dimension, but that also contains any of its members or descendants if the user has set an appropriate flag. Thus, the system maintains information on the dimensions and the dimension children. Moreover, the system can also include results from parents of a term when requested to do so. That is, the system can go up or down a hierarchical tree depending on the type of search the user desires to use.

[0082] Referring now to Figure 7B, the Repository\_Map table 718 of Figure 7A is shown in detail and includes records (rows) 742 of the cube dimensional metadata recorded by the wizard of Figure 6B. The records 742 detail the OLAP dimensions and corresponding tables and columns where the details of dimensional members involved in any discovered exceptions and patterns is recorded.

-18-

[0083] Referring now to Figure 7C, the constraints table 720 of Figure 7A is shown in detail and includes columns 744 and 746 of cube security metadata for all users or a specific user, respectively. The wizard automatically records this information during the results submission step shown here being imported from an OLAP database table 748 to the constraint table 720 via importation arrow 750. The PW\_Constraint\_Definition field describes information access privileges for any user belonging to any particular role or user group. These constraints are automatically and transparently applied when any authorized user interacts with the data repository.

[0084] Referring now to Figures 7D and 7E, two alternate schema to the facts table 716 are shown. Looking at Figure 7D, one alternate schema, generally 752, replaces multiple dimension specific tables with one dimensional table where details related to multiple dimension are recorded in multiple records sharing the same ID. Looking at Figure 7E, the second alternate schema, generally 754, includes a single large de-normalized table which stores content and context information all in one table. Figure 7E also shows the corresponding OLAP database schema 756.

[0085] The two alternate schemas are preferably applied in cases with lower dimensionality and in cases where the underlying RDBMS software allows large record size allocations. Regardless of the exact schema implemented, the schema provide similar functionality transparency to the end user. Also, the exact schema implemented can include customized functionality, unique to the customer.

[0086] Referring now to Figures 7F-I, metadata associated with other tables of Figure 7A are shown. Looking at Figure 7F, the Nugget Facts table 716 of Figure 7A is shown in detail and includes a field name column 758 and a corresponding description column 760, where the entries 762 in the description column 760 describe or give analytical and context information about the particular Information Nugget. The nugget fact table 716 stores the Analytical information content of the Information Nugget, while the various Dimensional Tables (depending on the schema) record the Dimensional context of the Information Nugget. Each ID key uniquely identifies the combination of Analytical content and Dimensional context.

[0087] Looking at Figure 7G, the Issues table 724 of Figure 7A is shown in detail and includes a field name column 764 and a corresponding description column 766, where the entries 768 in the description column 766 define or describe each field in the field name

column 764. Subjective issues, which generally span multiple Information Nuggets, may be human generated and/or automatically generated and are based on various classification and clustering or segmentation techniques. Information Issues are designed to deliver higher level knowledge derived from individual Information Nuggets or collections of individual Information Nuggets.

[0088] Looking at Figure 7H, the Batch Tasks table 722 of Figure 7A is shown in detail and includes a field name column 770 and a corresponding description column 772, where the entries 774 in the description column 772 define or describe each field in the field name column 770. Evolved processes that publish potential Information Nuggets to the Repository. The contents of the repository include distilled informational units and reflect directly on what is important to the business subject at any given time.

[0089] Looking at Figure 7I, a preferred embodiment of a Summary table 776 of this invention is shown in detail and includes a field name column 778 and a corresponding description column 780, where the entries 782 in the description column 780 define or describe each field in the field name column 778. is one of many ways in which the contents of Wisdom repository can be summarized to derive secondary intelligence on subjects of interest on a periodic basis and allows for pursuing changing informational and exceptional subjects overtime.

#### **Wisdom Repository Interactions**

[0090] Referring now to Figure 8A, an image, generally 800, is shown, which illustrates an integrated and easy means of interacting with the repository. The image 800 includes a results window 802 listing results 804 from a data mining routine. To invoke the repository wizard of this invention, the user simply right click on a result 806 that the user want to deposit in the repository of this invention. The right click operation results in the display of a selection window 808 which includes a list 810 of actions one of which 812 activates the repository wizard which when activated pulls down another window 814 which lists actions items 816. The user selects author/review information nugget action item 818, which invokes the wizard interface window 820 described below. When the wizard is activated, it automatically reads the context of the selected result 806 and retrieves any existing information contained in the repository that may further explain or provide insights related to the selected result 806 (exception or pattern), thus the information is readily shared without the need for explicit

-20-

searching *etc* and the user can easily study existing information and make a decision about Authoring Information Nuggets or simply providing Information Reviews and/or Pursues ratings. Existing information is thus readily available without the need for explicit searching improving information utilization and reducing the entry of duplicate or redundant information into the system.

[0091] Referring now to Figure 8B, the interface window 820 is shown, which is a simple interface that records objective, subjective commentary, reference documents, presentations, or the like provided by user. The user does not need to provide any other details to catalog the Information Nugget. The unique ID is automatically generated using the Dimensional and Analytical context of the selected result 806. The interface 820 includes two windows 822 and 824. The window 822 includes a score box 826 with up and down arrows 828, where the score is the user's objective rating to the relevance of the selected result 806 (data exception, anomaly or pattern). The window 822 also includes a subjective commentary entry box 830, where the user enters any subjective information relating to the selected result 806. The window 822 also include a save button 832 and a close button 834, where the save button 832 allows the results 806 to be submitted to the repository and the close button 834 allows the user to exit the submission process. The window 824 lists any existing information nuggets that are the same or similar to the result 806. The window 824 includes an Nugget Recency area 836, which includes a Nugget ID display box 838 and a data authored display box 840. The window 824 also includes a Nugget Reviews area 842, which includes a times reviewed display box 844 with scroll button 846 and an average review score display box 848 with scroll button 850 and a read review button 852. The window 824 also includes a Nugget Pursues area 854, which includes a times pursued display box 856 and an average pursue score display box 858 and a rate nugget button 860. The window 824 also includes a Nugget Author area 862, which includes a authored by display box 864, a score display box 866, an authoring score display box 868 and a Author's comment display box 870. The window 824 also includes nugget scroll controls 872.

[0092] It is important to note that multiple Information Nuggets may be recorded under the same Dimensional and Analytical context information, such additions should be reserved for adding uniquely different perspective on a given finding. The wizard also prompts the user as shown in Figure 8C for user confirmation prior to entering the result 806 into the repository

-21-

to minimize unintentional information duplication. Looking at Figure 8C, a confirmation window 874 is shown and includes an add review button 876, an add new information button 878 and a cancel button 880. The add review button 876 allows the user to simply add a review to an existing Information Nugget without creating a new nugget minimizing redundancy in the repository. The add new information button 878 actually creates a new Information Nugget from the result 806; while the cancel button 880 allows the user to exit the submission process.

[0093] As soon as an Information Nugget is submitted and stored or persisted in the Repository, it is available for review to all users logged on to the repository through the various Information Access channels described herein.

[0094] Referring now to Figure 8D, an interface window 882 is shown that is activated by pressing the read review button 852 of the window 824 of Figure 8B. The window 882 includes the review recency area 836 and the nugget pursues area 854 and a nugget reviewer area 884. Each review can be viewed in a view window 886, which shows the review score and reviewer's comments. The interface 882 provides an easy mechanism for reviewing records of objective and subjective commentary provided by users during the review of existing Information Nuggets. Multiple reviews may be provided by various users for any Information Nugget. As soon as an Information Nugget Review is persisted, it is available to all users how are connected to the repository.

[0095] Referring now to Figure 8E, a window 888 is shown that provides a simple interface that records an objective score related to the information value of an Information Nugget or an Information Nugget Review. The window 888 is invoked by activating the rate nugget button 860 and includes a score box 890 with controls 892 and an OK button 894 and a close button 896. Multiple ratings may be provided by multiple users. The main idea behind the objective ratings is to objectively score the Information Nuggets and Information Nugget Reviews for Information Value, thus making individual Information Nugget or Reviews more or less prominent in any subsequent query, search, pursue, or review.

[0096] Referring now to Figures 9A-B, sample entries made into some of the repository tables when a result is submitted to the repository are shown. The information is stored in the repository tables in an open and self documented format that is easily integrated into any custom application for further use. The information can also be easily queried using ure

-22-

friendly interfaces that hide much of the query complexities and present the information to the end-user in an easy to understand format. Additional columns can be added to the Information Nuggets Fact table to persist formatted results such as GIF images, XML and PDF documents, or the like adding additional information content to the stored Information Nuggets.

[0097] Looking at Figure 9A, a sample Nugget Fact table, generally 900, is shown, which displays the table information of the Information Nugget 806 of Figures 8A-B. The table 900 includes two columns 902 and 904. The column 902 includes field names 906a-x, while column 904 includes values 908a-x associated with the names 906a-x. The PW\_Wisdom\_Type field 906d and the PW\_Wisdom\_Class field 906e of the Nugget Fact table 900 define the Analytical context of the Information Nugget 806. Fields 906f-k and their corresponding values 908f-k include source database information. Fields 906l-x and their corresponding values 908l-x include Information Nugget information specific to the repository.

[0098] Looking at Figure 9B, sample dimensional tables 910, 912, 914, and 916 are shown to include corresponding entries made into these table during the storage of the result 806 as an Information Nugget in the repository. The table 910 includes information about the Customer dimension; table 912 includes information about the Education dimension; table 914 includes information about the Gender dimension; and table 916 includes information about the Measures dimension. Each field of the dimension specific tables correspond to the level of the OLAP dimension and are arranged in the table in the order of their dimensional cardinality from left (0 left most) to right, the fields persist details of members belong to the corresponding dimension level. The fields represent details of an Information Nugget Member that belongs to a corresponding dimension level. Such a design permits retrieving the Information Nuggets through the use of "LevelMembers", "Ascendants" and "Descendants" of a member like OLAP functionality emulated through SQL Constructs. All dimensional table share the common unique Information Nugget ID so that the dimensional context of any Information Nugget is uniquely recorded across the dimensional table. Moreover, the ID is a unique for each Information Nugget in the specified dimensional context and analytical context.

[0099] Looking at Figure 9C, an overview of a preferred information catalog, generally 920,



-23-

of this invention is shown that further illustrates the cataloging of Information Nuggets within the repository. The overview 920 includes an Information Nugget section 922 including multiple Information Nuggets 924. These Information Nuggets 924 are associated with a dimensional context 926, an analytical context 928 and optionally any other context 930 that a customer wishes to add. Multiple Information Nuggets may exist in the repository for a given combination of contexts; however, each Information Nugget associated with a given combination of contexts are preferably distinct due to their associated objective and/or subjective commentary. Thus, the user only focuses on recording subjective matter and the Information Nugget is uniquely recorded and is readily searchable based on the attributes of a context within the repository.

#### **Information Delivery**

[00100] The system provides for simple, intuitive and transparent information query interfaces to shield the user from many of the complexities in searching for the relevant Information Nuggets. Referring now to Figure 10A, a Drag and Drop query interface, generally 1000, is shown, which includes a cube window 1002, an repository search window 1004 and an Information Nugget window 1006. The cube window 1002 includes cube dimensional and measures levels 1008, with associated expand/contract buttons 1010. The search window 1004 includes a drop box 1012, where a cube dimension or dimension member such as high school degree 1014 can be dragged and dropped into the box 1012. Once in the box 1012, the user can submit a search to the repository by pushing a search button 1016. The window 1004 also includes a results section 1018, where the all Information Nuggets satisfying the search criteria are displayed. If a given result is selected, then the window 1006 is activated, which the wizard of Figures 8A-E.

[00101] The Information Nuggets presented in the analysis pane 1018 of this invention are organized by Nugget attributes such as Exception or Rule Type and can be sorted based on a variety of quantitative measures. User can simply right click on the retrieved Information Nugget and read details including Reviews and Pursue ratings, the user can also introduce new Information Reviews and Information Pursues or simply drag the Information Nugget onto the right in Cross-Tab pane and retrieve latest values for the Information Nugget and so forth.

[0102] Looking at Figure 10B, the SQL script 1020 that is automatically generated for the above user search is shown. Filters such as user security constraints *etc.* are automatically

-24-

applied to the SQL (as the interface reads such constraints from the user profile at login in time). Since the Nugget repository is designed to have specific nomenclature, custom interfaces can be easily programmed to generate SQL statements that can interact with the Nugget repository. Looking at Figure 10C, a flowchart 1022 evidencing the user query of Information Nuggets repository is shown. The flowchart 1022 includes an search step 1024 where the user query is submitted to the repository, which checks for information nuggets in a conditional step 1026. The conditional step 1026 checks the repository to determine if any Information Nugget satisfy the search criteria. The conditional step 1026 provides for three different conditions. First, if Information Nuggets exist and the user wants to add objective ratings and/or subjective comment on any of the existing Information Nuggets, then control is transferred along a YES branch 1028 to an information nugget review step 1030 which produces an information review document 1032 that is added to the repository. Second, if Information Nuggets exist and the user wants to objectively rate the importance of an Information Nugget, then control is transferred along a YES branch 1034 to an information nugget pursue step 1036 which produces an information pursue document 1038 that is added to the repository. Finally, if Information Nuggets do not exist that satisfy the query criteria, then control is transferred along a NO branch 1040 which can optionally includes a record search step 1042. Subsequent automated and user driven exception and pattern driven routines can then focus on these search parameters to identify any possible Information Nuggets.

[0103] As the repository gets populated with numerous nuggets, an important challenge is to present the user with the most relevant Information first. The system accomplishes this by utilizing the following repository information about the Information Nuggets: (1) Reviews and Pursue Ratings provided by the knowledge worker community, (2) User's explicit information preferences (maintained in the Wisdom repository centrally), which is recorded through the "My Information Profile" interface shown in Figure 11, (3) User's explicit member selections as shown in Figure 10A; and (4) Suggestive Approach— user's search and review activity is recorded overtime such that intelligence related to what may be interesting is based on his or her historical activity and/or lack of (in particular subject areas).

[0104] Referring now to Figure 11, a simple Interface, generally 1100, which records user informational preference profiles to minimize the problem of Information Glut or information

-25-

overload. The interface 1100 includes an Information Filters tab 1102, a Author Filters tab 1104, a Review Filters tab 1106 and a General Preferences tab 1108. The Information Filters tab 1102 includes an Information Type area 1110, which includes exceptions filters 1112 and pattern rule filters 1114. The Information Filters tab 1102 also includes an Information Recency area 1116, which includes time filters 1118. The Information Filters tab 1102 also includes an Information Limit area 1120, which includes limit filters 1122. The Information Filters tab 1102 also includes an Information Relevance area 1124, which includes relevancy filters 1126. The interface 1100 also includes a reset all button 1128, a save button 1130, a cancel button 1132 and a close button 1134. The user can setup Time, quantitative and qualitative filters besides setting up a subscription profile, wherein user can pursue Information Nuggets and Reviews authored by a select list of people. These filters can be applied in variety of ways.

[0105] Once setup, the user's information profile is recorded in the repository, it is available for future activity and across various information channels. The Information Profile based filters are transparently applied in the query as described in Figure 10B. Additionally, where meaningful, a decision making workflow and "Push n Pull" mode for interacting with the Information Nuggets can be easily built around the repository.

[0106] Referring now to Figures 12A-G, a preferred web-based interface, generally 1200, built around the Repository is shown, which enables easy and customized information delivery, where the self-documented and open architecture of the repository allows for such custom information delivery and integration needs. Figure 12A shows the home page of the Custom Interface 1200, which includes four search options: search by Nugget ID 1202, search by Discovery Process 1204, search by Information Issues 1206 and Ad-hoc Search 1208. The users can choose from these options to search for the Information Nuggets within the repository. Figure 12B illustrates an interface screen 1210 which presents users with a list of Evolved or Batched Data Mining processes and includes a batch discovery process area 1212, a process run data area 1214 and an Information Nugget area 1216. The user makes selection of a particular process and log activity for the specific process is presented. The user can select a process run and retrieve Information Nuggets belonging to that run. Additionally, an interface can present the changes between the Information Nuggets produced by the Batch process overtime, etc. Users may further Review or Pursue the Information Nuggets that are

-26-

presented as described above. Figure 12C illustrates an interface screen 1218 including an Information Issues area 1220 and an Information Nuggets area 1222. The users are first presented with the Information Issues in the area 1220 contained in the repository and when the user makes a selection of the Information Issue – all the related Information Nuggets in the area 1222 are presented to the users. Figure 12D illustrates an ad-hoc query interface 1224, which allows users to query Information Nuggets based on various filters and attributes including free-text search of Information Nuggets.

[0107] In addition to these interfaces, since the repository is knowledgeable of each users past interactions, it can simply present the Information Nuggets that the user will be potentially interesting to the user, as soon as the user logs in to the repository. Also, the information content of repository may be presented in a hierarchical format similar to the OLAP dimension navigator. Thus the user knows what exists in the repository and easy access to the members involved in Information Nugget.

[0108] Referring now to Figures 12E-G, another web based interface, generally 1260, which supports English language query processing is shown. Figure 12E shows the interface 1260 where user can simply types in a question in English in a text box 1262 and retrieve related Information Nuggets.

[0109] Once the base question is parsed, clarifications may be sought along with recommendations on the selection of additional dimensional members, which would fetch Information Nuggets that are of potentially greater interest. Figure 12E an search results screen 1264 is shown which is seeking clarification from the user related to the keyword CA in items 1266 above the horizontal line 1268. Dimensional Members that together with this keyword are associated with interesting results are suggested for user selection (as presented below the Horizontal line) such that specific interesting results may be presented quickly.

[0110] Figure 12F next, shows web based rendering 1270 of a results retrieved from the repository that satisfied the search criteria. The report elements can be previously persisted or dynamically re-generated based on the Information Nugget metadata. If no previous result exists for the original question of interest, the keywords can be recorded to the a Search Table and the scheduled/On Demand Data Mining tasks can include these members in data exploration to yield relevant information.

#### **Information Summarization and Integration**

-27-

[0111] The attributes related to the Information Nuggets can be summarized and presented in a variety of ways so as to give a holistic picture of the contents of the Nugget repository. Referring now to Figure 13, a sample interface 1300 is shown, where the entire contents of repository can be organized in a Dimensional Tree like format 1302. Here Distinct members from various fields in the Dimensional Tables are organized in an hierarchical format and their summarized attributes such as References in Repository Information Nuggets, Average Review and Pursue Score (overtime) etc. can be presented. Secondary intelligence, such as summaries, information clusters and any other form of abstractions derived from the repository is of vital importance as it allows for quickly scanning through and getting the larger picture of issues and opportunities that are detailed in the form of Information Nuggets, Information Issues and other custom abstracted forms.

[0112] Any Data Mining algorithm can be integrated into the database of this invention, meaning the dimensional member contained in the repository can be explicitly explored or ignored by the Data Mining algorithms as they search the base data cube for anomalies and patterns, Dimensional member associations recorded by certain algorithms such as Association and Market Basket Analysis can be used as hints in *a priori* candidate generation improving the performance of these algorithms in subsequent runs *etc.* Figure 14 shows an example of Information Integration 1400 wherein Basket Analysis Data mining algorithm task 1402 is customized per the contents of Nugget repository 1404 to generate a result 1406. Data Mining algorithms or routines can be made repository or system aware, meaning the dimensional member contained in the repository can be explicitly explored or ignored by the Data Mining algorithms as they search the base data cube for anomalies and patterns. Dimensional member associations recorded by certain algorithms such as Association and Market Basket Analysis can be used as hints in *apriori* candidate generation improving the performance of these algorithms in subsequent runs *etc.*

### Conclusions

[0113] An open, extensible, self-documented and self-cataloging knowledgebase interacted through simple, flexible and extensible collaborative decision making and information delivery, integration and abstraction mechanics greatly improve the creation of and value derived from corporate knowledge base. The unique repository schema and information persistence along with the qualitative and quantitative information metrics allow in enhanced

-28-

information query processing, while minimizing the traditional problems associated with "Information Glut", "Information Islands" and "Information Cataloging". The systems server based model is transparent and its role based user interfaces permit effective leveraging of collective knowledge, experiences and techniques of information workers and extends its value to a broader base of company's information consumer community, who can be completely oblivious to the information/knowledge creating analytics and processes and are simply offered the fruits (results) of analysis.

[0114] All references cited herein are incorporated by reference. While this invention has been described fully and completely, it should be understood that, within the scope of the appended claims, the invention may be practiced otherwise than as specifically described. Although the invention has been disclosed with reference to its preferred embodiments, from reading this description those of skill in the art may appreciate changes and modification that may be made which do not depart from the scope and spirit of the invention as described above and claimed hereafter.

-29-

**CLAIMS**

We claim:

1. A database comprising information nuggets including sufficient information to identify a source of each information nugget, object human ratings of each information nugget, subjective human commentary associated with each information nugget, where the information nuggets comprise data exceptions, metaexceptions, and/or patterns and where the database is adapted to receive, store, and retain results from human and/or machine data mining activities so that information or knowledge can be derived and/or created therefrom.
2. A database structured to store non-redundant results from human and/or machine data mining activities, where the database comprises information sufficient to reconstruct each stored result, information on each query that lead to each stored result and a structure sufficient to facilitate efficient retrieval, manipulation, integration, and distribution of the stored results.
3. A method for storing results from human and/or machine derived data mining results including the step of receiving a result from a human and/or machine data mining activity, comparing the received result to previously stored results, storing the received result if the received result is new or updating a previously stored result if the received result is an augmentation, addition, correction or update of the previously stored result and if the received result causes a change in the database notifying a user group of the change.
4. The method of claims 3, further comprising the step of adding objective and/or subjective commentary, analysis, and/or rating.
5. The method of claims 3, further comprising the step of reviewing the results to minimize redundant results and adding further objective and/or subjective commentary, analysis, and/or rating by a group of specialized users.

-30-

6. The method of claims 3, further comprising the step of integrating the results into summary, collective, abstracted or compound results.
7. The method of claims 3, further comprising the step of running data mining routines against the results to derive higher order data patterns, trends or exceptions.
8. The method of claim 7, wherein the running step occur either continuously or periodically.
9. The method of claim 6, wherein the running step occur either continuously or periodically.
10. An interactive method for acquiring knowledge from data mining results including the steps of receiving input either manually via direct human entry into a knowledge database or automatically from participating data mining routines, storing the results in a temporary database area, screening the result against existing result entries, marking the results as either new, related or redundant, authoring the temporary results with objective and/or subjective commentary, analysis and/or rating, submitting the authored result to the database for permanent storage, reviewing the results by a special group of users to add expert subjective and/or objective commentary, analysis and/or ratings, forming higher level results from the stored results, storing the higher level results in the database, removing redundant results from the database, presenting relevant results to specific user groups, running data mining routines against the results and integrating the results into patterns and trends and repeating the above steps on a continuous or periodic basis.
11. A knowledge storage and management system comprising an input interface that allows results from one or more data mining routine running on one or more data sources or from one or more human data mining activities to be screened for inclusion in an knowledge



-31-

acquisition database which stores the results of the data mining activities with sufficient context information to recreate the data mining process and the source data structure into a knowledge repository and management database; a knowledge review interface and pursue interface that permit the entry of human objective and/or subjective commentary to add additional informational content to the member of the database; specially tailored data mining routines that run against the database to produce higher level knowledge and informational content; and an action interface information that provides action items to the database relating to the use of the knowledge or information in the system.

1/23

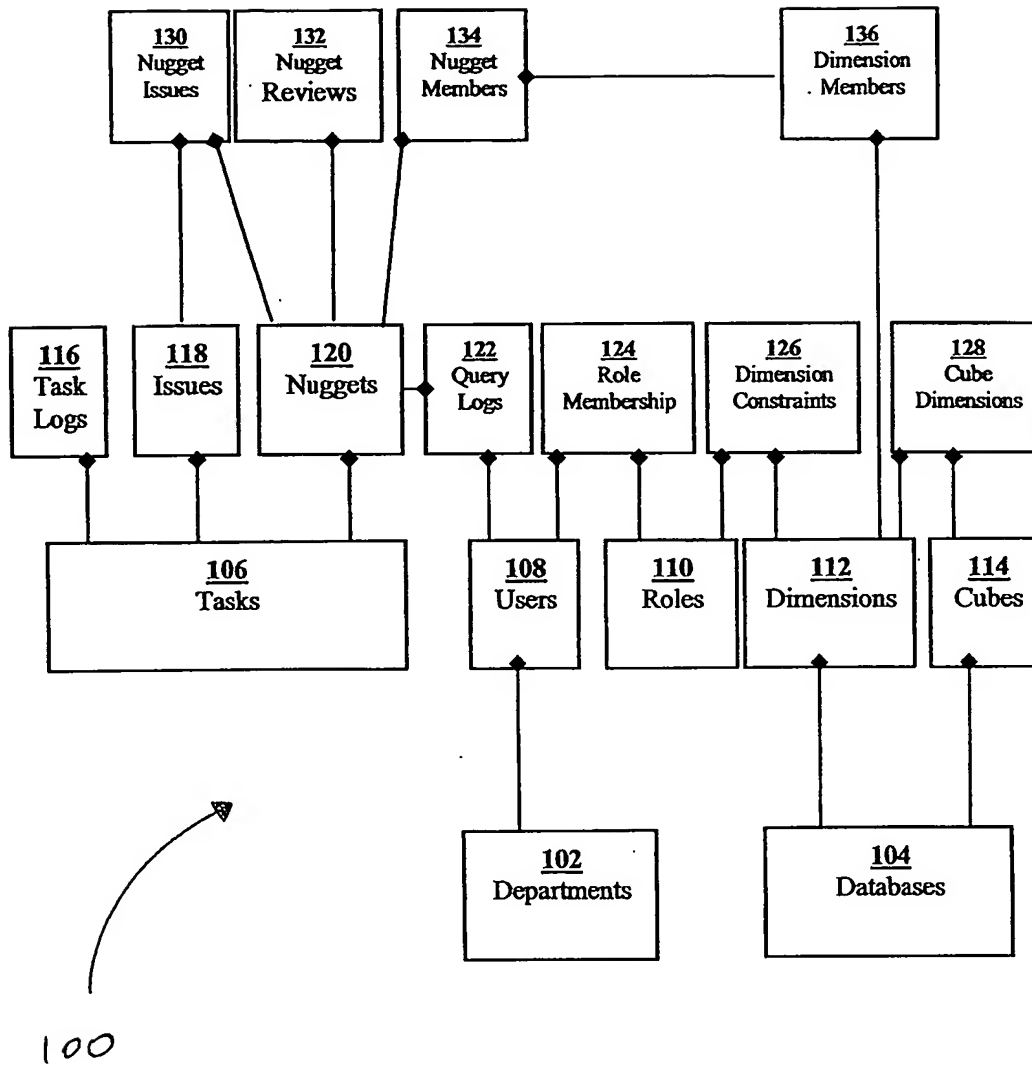


FIG. 1

2/23

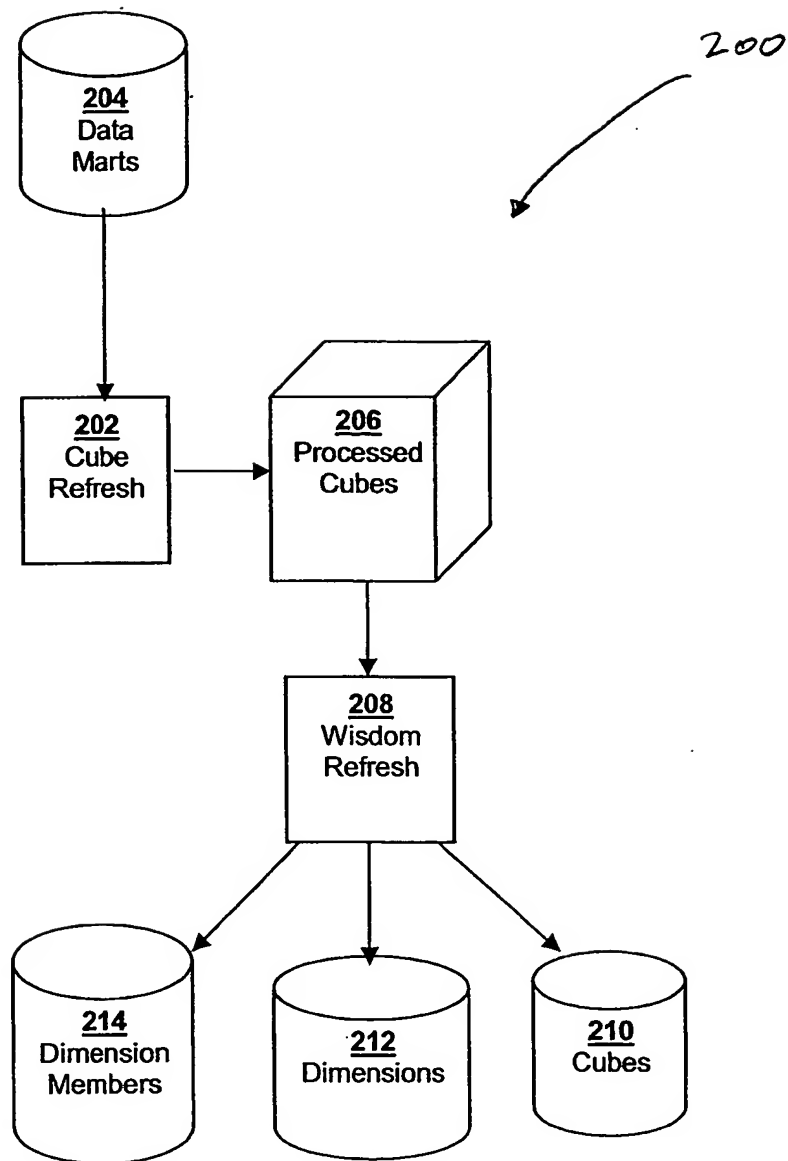


FIG. 2

3/23

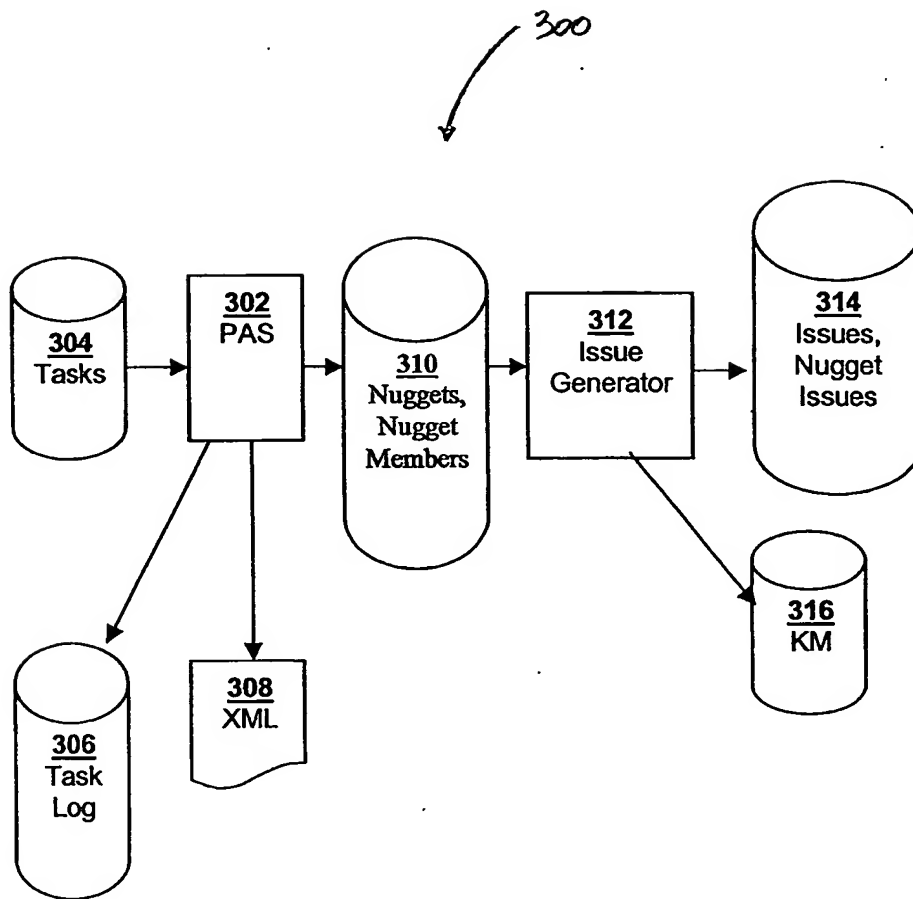


FIG. 3

4/23

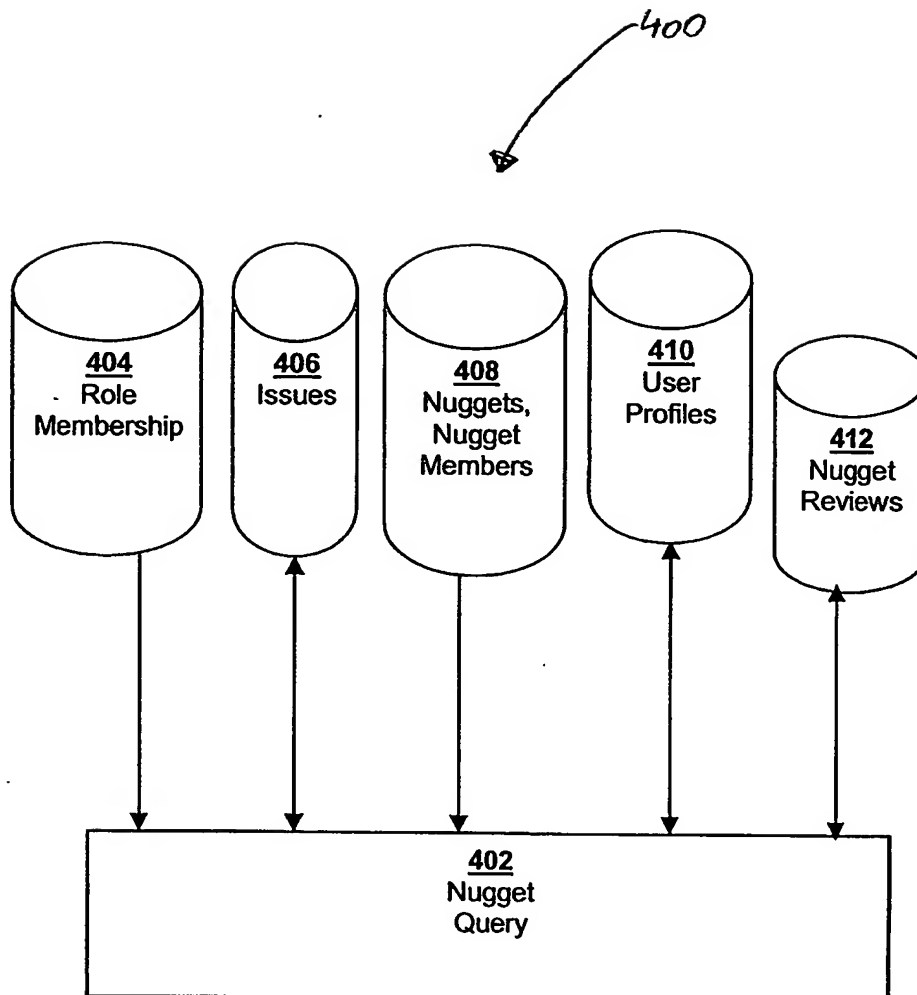


FIG. 4

5/23

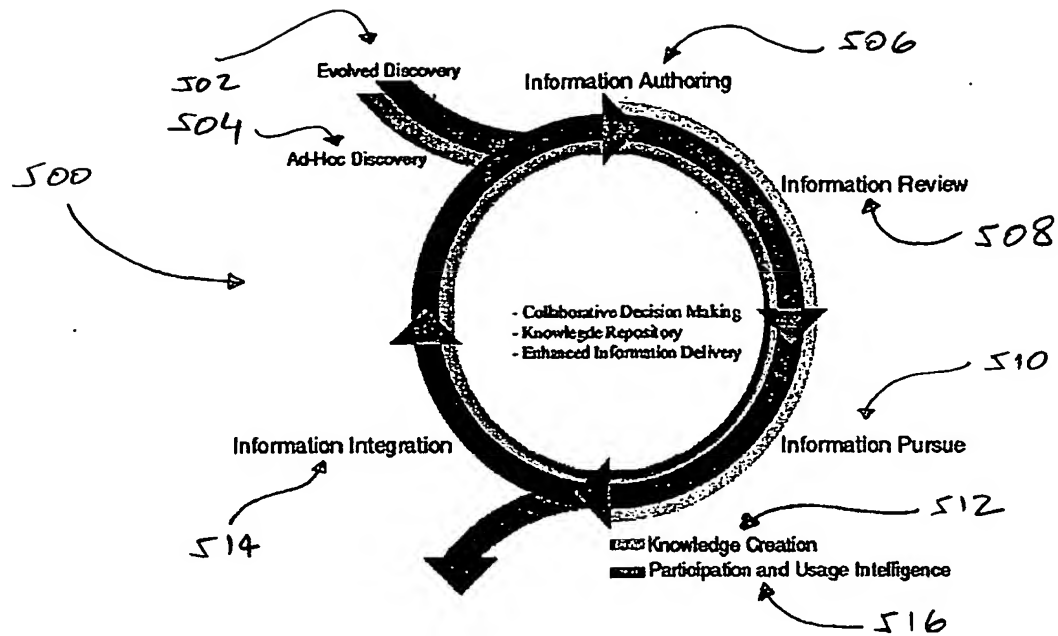


FIG. 5A

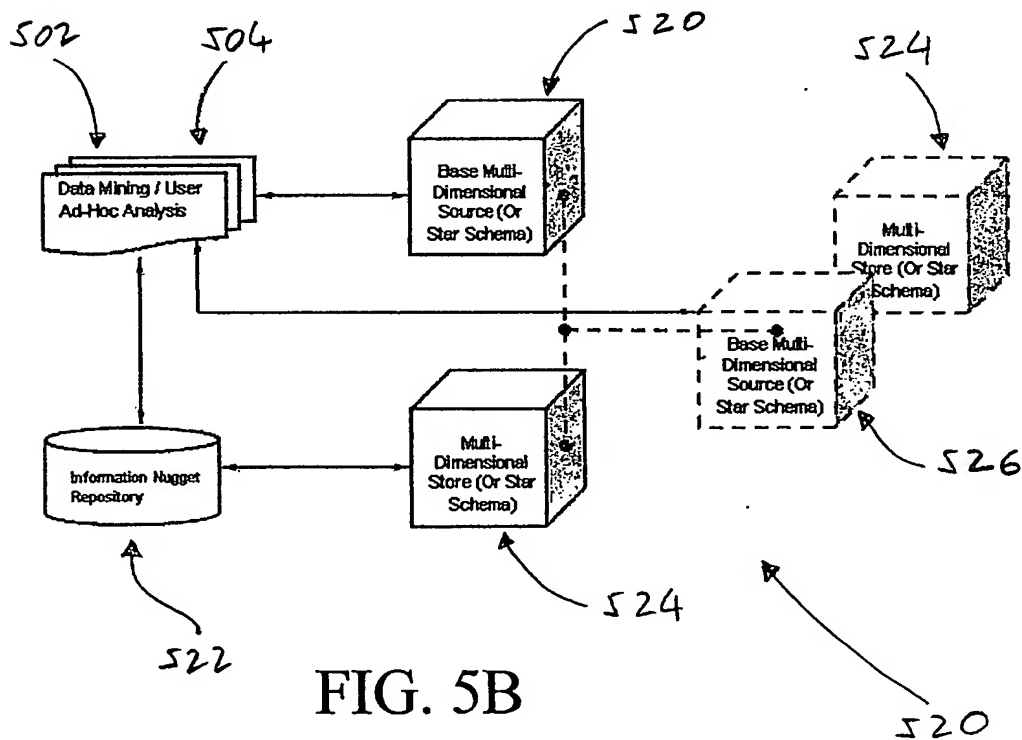


FIG. 5B

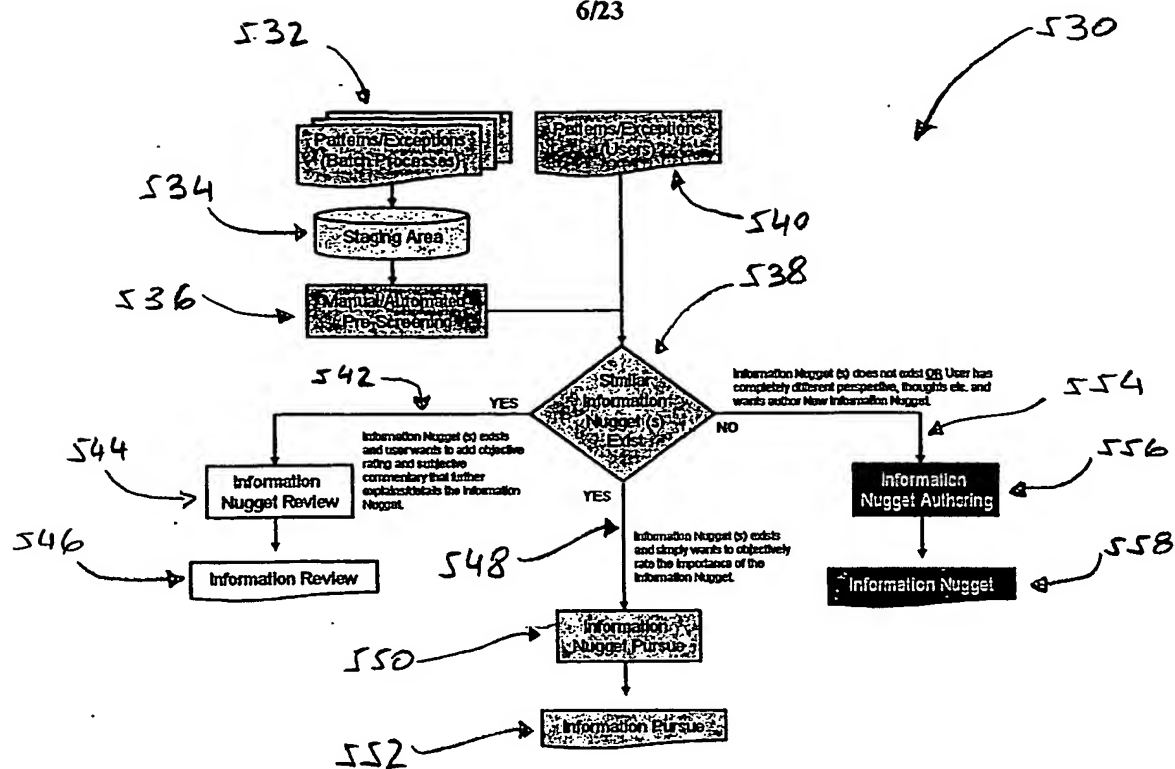
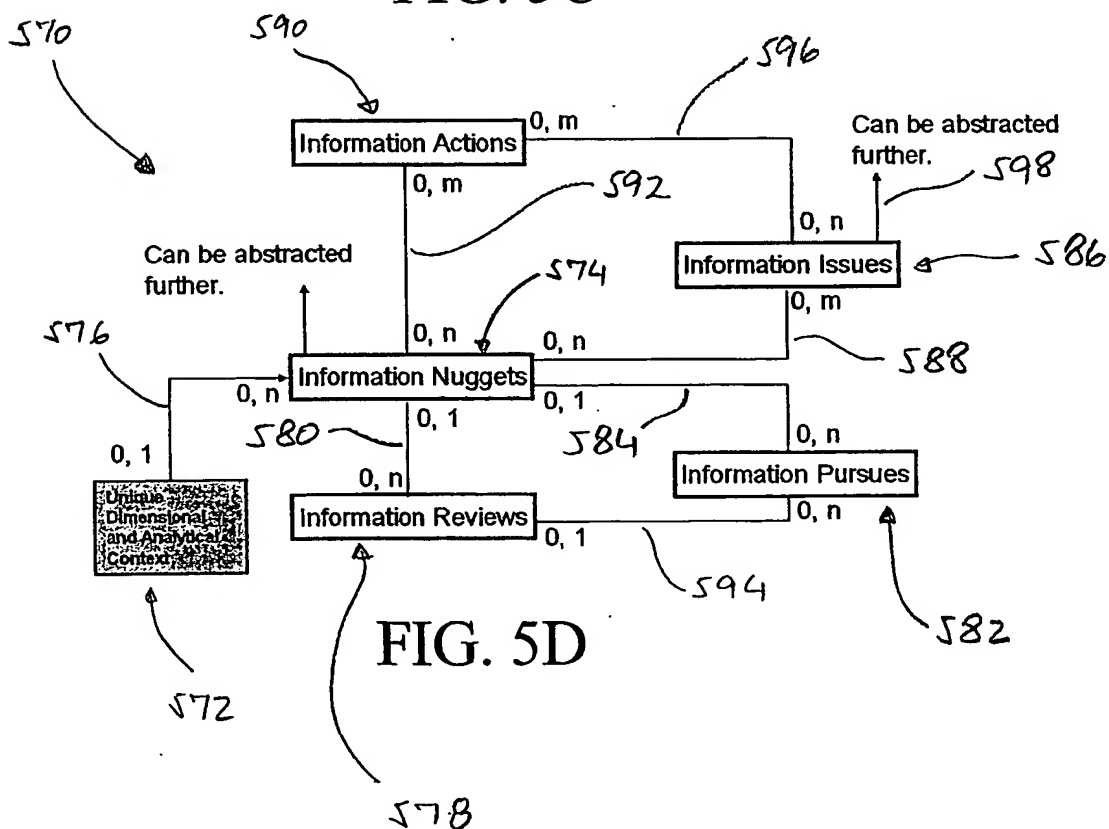


FIG. 5C



7/23

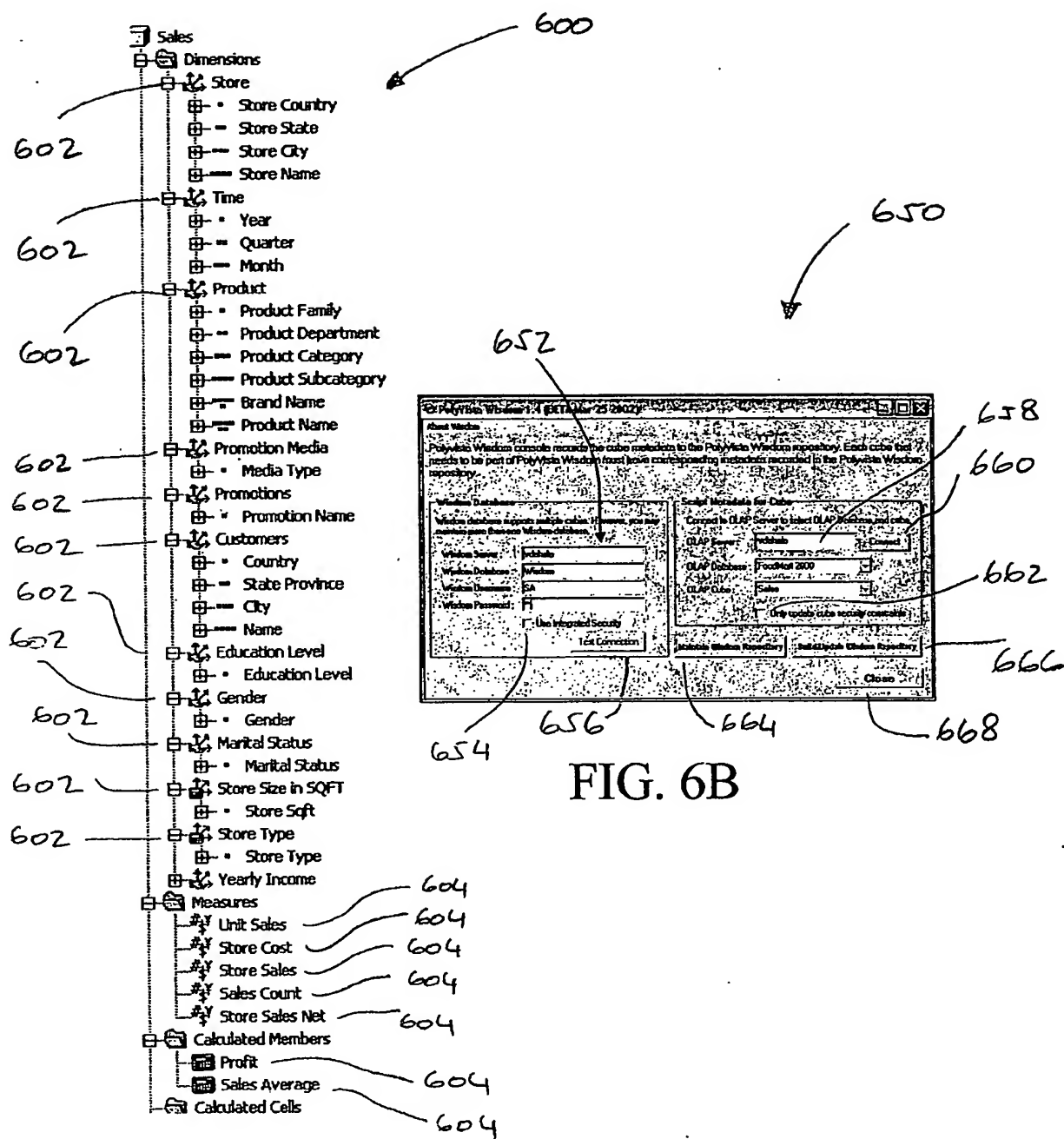


FIG. 6B

FIG. 6A



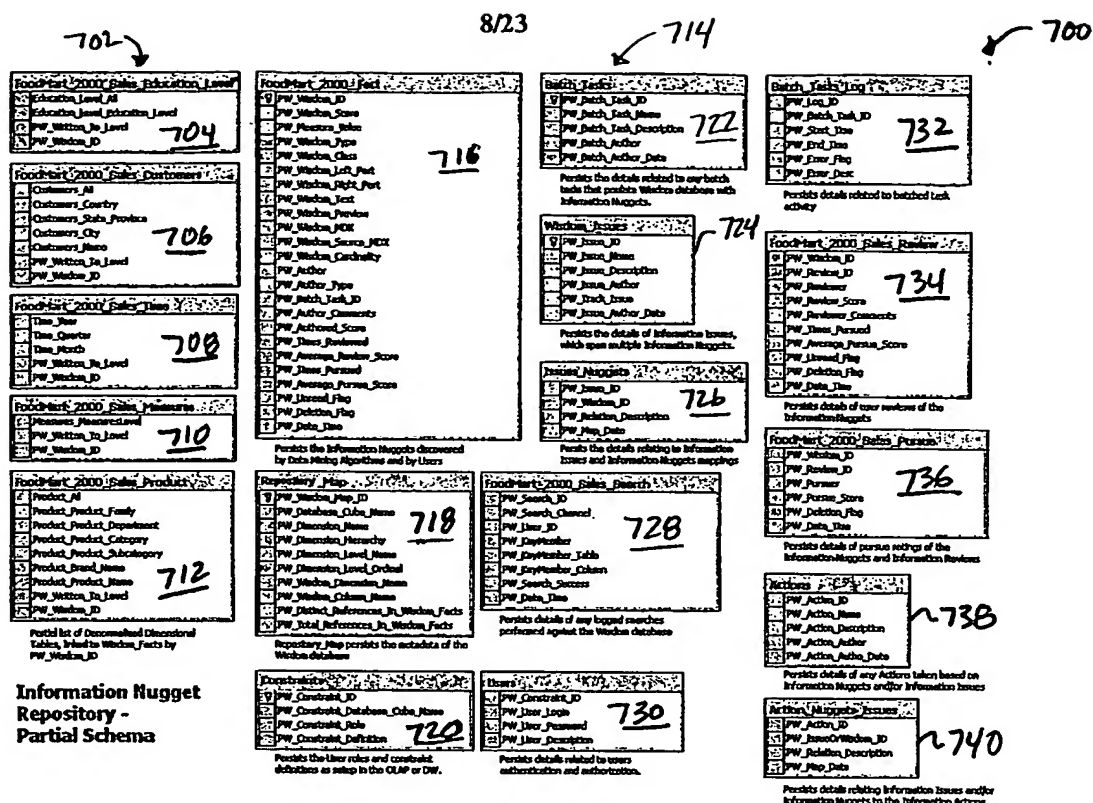


FIG. 7A

[illegible]

FIG. 7B

**9/23**

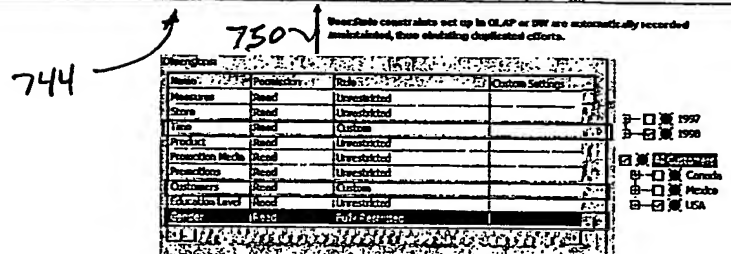
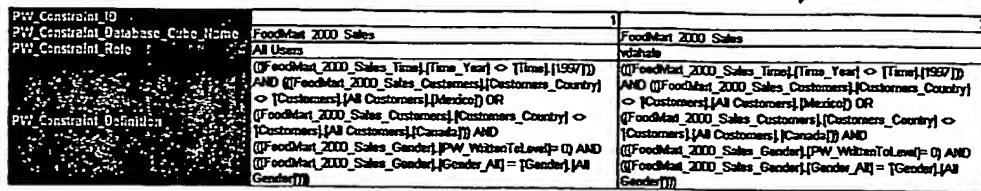


FIG. 7C

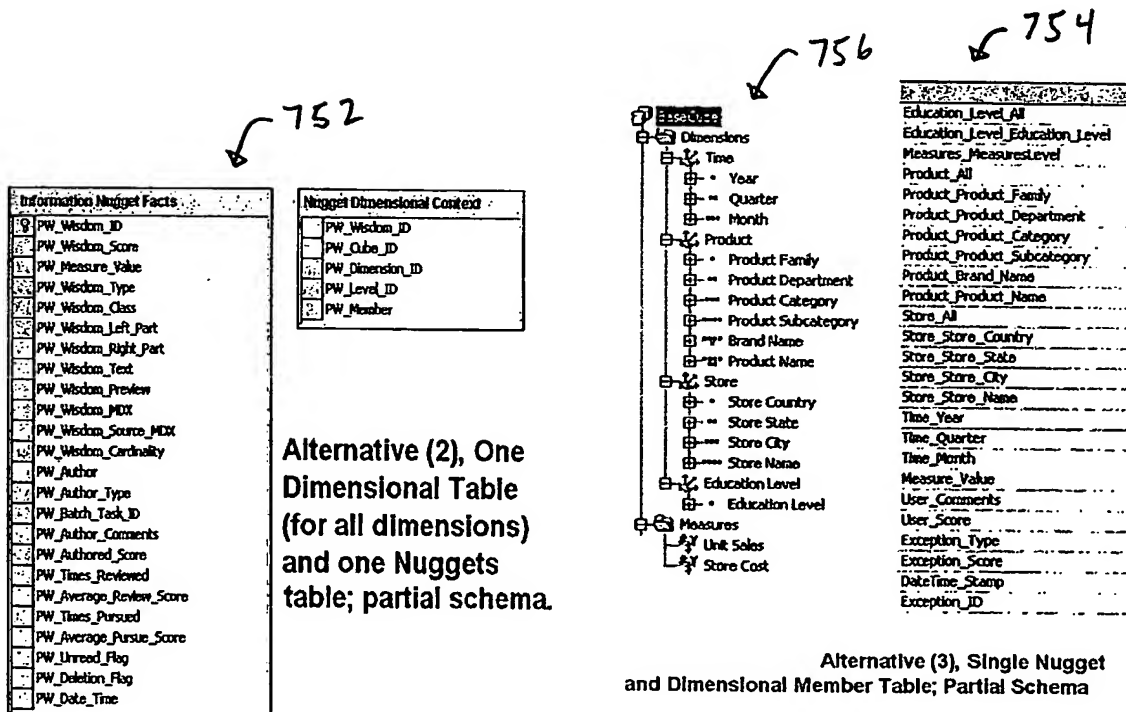


FIG. 7D

FIG. 7E

10/23

Field Name	Description
PW Wisdom ID	Uniquely identifies the Information Nugget (Exception or Pattern)
PW Wisdom Score	Composite Score representing Statistical Significance (where applicable)
PW Measure Value	Measure data value for the Information Nugget
PW Wisdom Type	Attribute of Information Nugget (Deviation Analysis, Affinity Analysis, Clustering, etc.)
PW Wisdom Class	Attribute of Information Nugget (Exception, Pattern, etc.)
PW Wisdom Left Part	Information Nugget Left Part (Where Applicable)
PW Wisdom Right Part	Information Nugget Right Part (Where Applicable)
PW Wisdom Text	Information Nugget Text - Complete Description of the exception or pattern
PW Wisdom Preview	Cached GIF, Crosstabs, English Text etc. of the Information Nugget for fast reporting
PW Wisdom MDX	SQL or MDX statements to retrieve information related to the Information Nugget from the OLAP cube or DW
PW Wisdom Source MDX	SQL or MDX statements to retrieve information related to the context of the Information Nugget from the OLAP cube or DW
PW Wisdom Cardinality	Dimensional Member cardinality involved in the Information Nugget
PW Author	Author of the Information Nugget
PW Author Type	Automated Evolved Tasks or User
PW Batch Task ID	Batch Task that may have introduced the information nugget.
PW Author Comments	Subjective Commentary related to the Information Nugget provided by the Author of the Nugget
PW Authored Score	Objective Score (on a scale of 1 to 10) related to the Information Nugget provided by the Author of the Nugget
PW Times Reviewed	Regularly updated score on how many Information Reviews exist for the Information Nugget
PW Average Review Score	Regularly updated score on average of Information Reviews scores for the Information Nugget
PW Times Pursued	Regularly updated score on how many Information Previews (Pursued and rated) exist for the Information Nugget
PW Average Pursue Score	Regularly updated score on average of Information Pursue scores for the Information Nugget
PW Unread Flag	Indicates whether the Information Nugget has been reviewed or Pursued yet
PW Deletion Flag	Indicates whether the Information Nugget has been marked for deletion
PW Date Time	Date when the Information Nugget was introduced in to the Wisdom repository

Facts (Nuggets) Table;  
Partial Schema

FIG. 7F

Field Name	Description
PW Issue ID	Uniquely identifies a group of Information Nuggets associated with a larger issue or having other characteristics
PW Issue Name	Name of the Information Issue
PW Issue Description	Description of the Information Issue
PW Issue Author	Author of the Information Issue
PW Track Issue	Indicates whether the issue is marked for tracking, in which case appropriate Evolved Tasks will examine the data for the exceptions and patterns covered by the issue on a periodic basis
PW Issue Author Date	Date when the Information issue was authored or revised

Issues Table; Partial Schema

FIG. 7G

Field Name	Description
PW Batch Task ID	Uniquely identifies Evolved Discovery Tasks, which publish potential information nuggets to Wisdom Repository
PW Batch Task Name	Name of the Evolved Discovery Task (Data Mining or any Statistical or Analytical Method)
PW Batch Task Description	Description of the Evolved Discovery Task
PW Batch Task Definition XML	Definition of the Evolved Discovery Task in XML format (PMML or PolyVista Format)
PW Status Flag	Indicates whether the task is active or inactive.
PW Schedule ID	Describes the Auto Run Schedule of this task
PW Batch Author	Name of the user who defined this task
PW Batch Author Date	Date the Task was Authored or Revised

Batch Tasks Table; Partial Definition

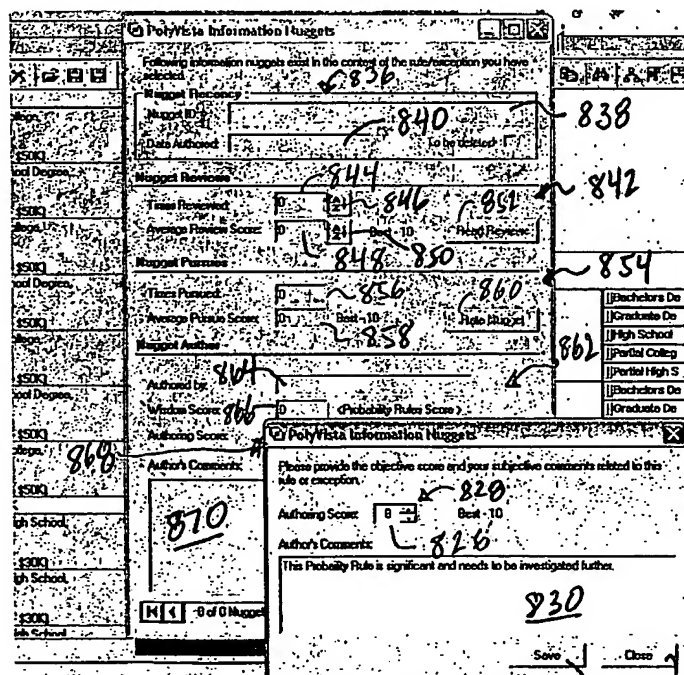
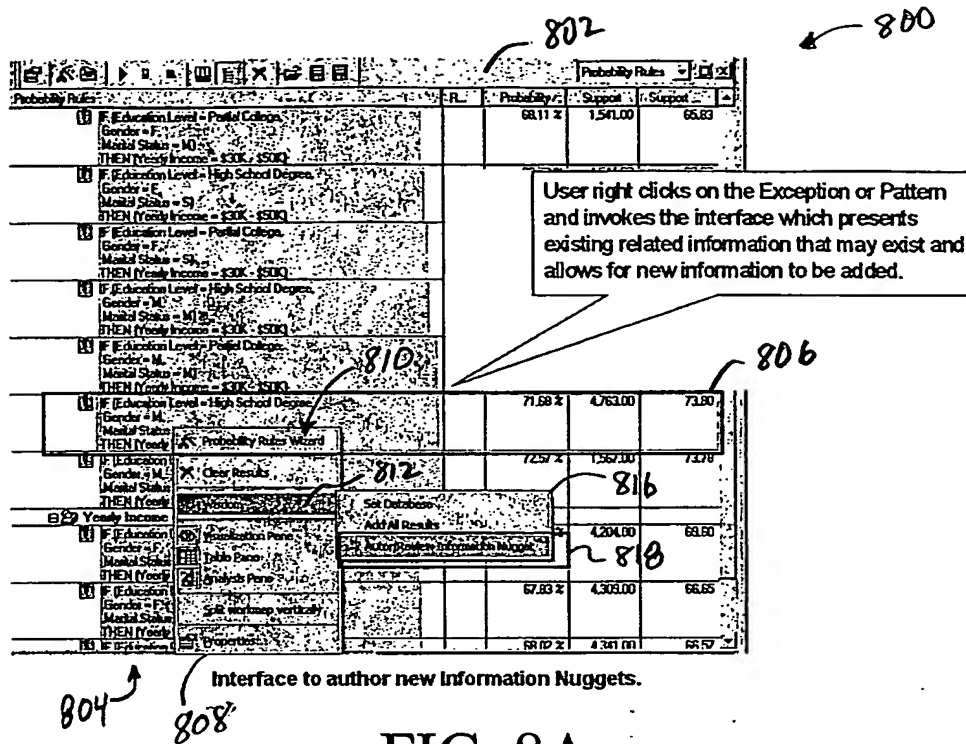
FIG. 7H

Field Name	Description
PW Wisdom Summary ID	Uniquely defines the Wisdom Summary Snapshot
PW Wisdom Cube Name	OLAP cube or DW which is the source of Information Nuggets
PW Wisdom Dimension Name	Dimension whose members are part of the Information Nuggets
PW Wisdom Column Name	Dimension Level whose members are part of the Information Nuggets
PW Positive References	No. of times the Dimensional member is involved in good exceptions and patterns
PW Negative References	No. of times the Dimensional member is involved in bad exceptions and patterns
PW Positive Effect	% Value of Positive References
PW Negative Effect	% Value of Negative References
PW Review Count	Number of times the Information Nugget is reviewed
PW Pursue Count	Number of times the Information Nugget is pursued
PW Search Count	Number of times the Dimensional Member is searched for
PW Action Count	Actionable event count resulting from the Information Nugget
Avg Review Score	Average Information Review Score
Avg Pursue Score	Average Information Pursue Score
PW Summary Updated	Timestamp of when the summary was last created/updated

Summary Table; Partial Schema

FIG. 7I

11/23



12/23

Previous information suggests death. Have you passed existing information suggests? You may add suggested that present uniquely new information. If you want to explain, contradict, elaborate existing information, adding evidence is recommended.

876 878 880

Add Previous Add New Information Cancel

**Wizard prompts for user confirmation to prevent addition of redundant information.**

FIG. 8C

882

836

854

884

824

852

886

Interface to review existing Information Nuggets.

FIG. 8D

13/23 888

890 892 894 896

Information Value: 5 On a scale of 1 to 10 (Best)

OK Close

860

Interference Information Nuggets Pursue

Please provide the rating of the Information Nugget or Review that you have just pursued. These ratings will help improve the relevance of Webdata search results.

Information Value: 5 On a scale of 1 to 10 (Best)

OK Close

Times Reviewed: 1 Best: 10

Average Review Score: 0 Best: 10

Read Reviews

Nugget Pursued:

Times Pursued: 1

Average Pursue Score: 0 Best: 10

Rate Nugget

Nugget Author:

Author: Veritastech Database

Webdata Score: 72

Probability Rules Score: 0

Authoring Score: 0 Best: 10

Author's Comments:

This Probability Rule is significant and needs to be investigated further.

1 of 1 Nuggets

Add Edit Delete Close

Interface to rate the relevance (Information Pursue Ratings) of existing Information Nuggets and Information Reviews.

FIG. 8E

14/23

Column Name	Column Value	
PW Wisdom ID	{576DD72C-A17A-4FB3-8A7F-00EB29013624}	908a
PW Wisdom Score 906b	71.68	908b
PW Measure Value 906c		908c
PW Wisdom Type 906d	16	908d
PW Wisdom Class 906e	2	908e
		908f
PW Wisdom Left Part	[Education Level].[All Education Level].[High School Degree],	
PW Wisdom Right Part	[Gender].[All Gender].[M].[Marital Status].[All Marital Status].[S]	
	[Yearly Income].[All Yearly Income].[\$30K - \$50K]	908g
PW Wisdom Text 906h	IF ( Education Level = High School Degree,Gender = M,Marital Status = S) THEN (Yearly Income = \$30K - \$50K)	908h
	[[Education Level].[All Education Level].[High School Degree],	
PW Wisdom Preview	[Gender].[All Gender].[M].[Marital Status].[All Marital Status].[S].[Yearly Income].[All Yearly Income].[\$30K - \$50K])	908i
	SELECT	
	[[Education Level].[All Education Level].[High School Degree]] ON	
PW Wisdom MDX	AXIS(0),	908j
	[[Gender].[All Gender].[M]] ON AXIS(1),	
	[[Marital Status].[All Marital Status].[S]] ON AXIS(2),	
	[[Yearly Income].[All Yearly Income].[\$30K - \$50K]] ON AXIS(3)	
	FROM [Sales] WHERE ([Measures].[Sales Count])	
	SELECT	
	[[Gender].[Gender].members] ON AXIS(0),	908k
	[[Marital Status].[Marital Status].members] ON AXIS(1),	
	[[Education Level].[Education Level].members] ON AXIS(2),	
	[[Yearly Income].[Yearly Income].members] ON AXIS(3)	
PW Wisdom Source MDX	FROM [Sales]	
	WHERE ([Measures].[Sales Count])	
PW Wisdom Cardinality	4	908l
PW Author 906m	Venkatesh Dahale	908m
PW Author Type 906n	U	908n
PW Batch Task ID 906o		908o
		908p
PW Author Comments	This Probaility Rule is significant and needs to be investigated further.	908p
PW Authored Score 906q	8	908q
PW Times Reviewed 906r	1	908r
PW Average Review Score 906s	8	908s
PW Times Pursued 906t	1	908t
PW Average Pursue Score	0	908u
PW Unread Flag 906v	0	908v
PW Deletion Flag 906w	0	908w
PW Date Time 906x	4/19/2002	908x

Sample Information Nugget Record in the Nuggets Fact Table

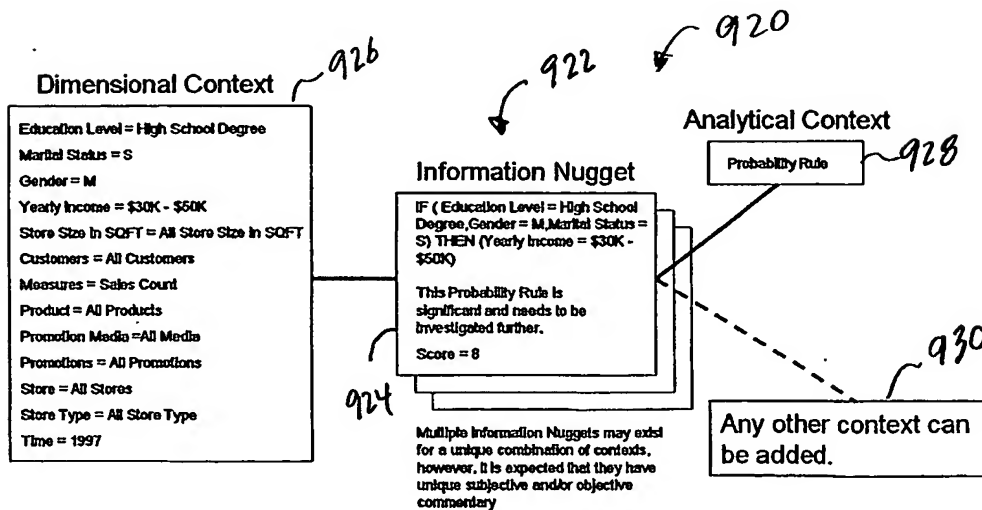
FIG. 9A

15/23

Column Name	Column Value
<b>Customer Dimension Table</b>	
Customers All	[Customers].[All Customers]
Customers Country	NA
Customers State Province	NA
Customers City	NA
Customers Name	NA
PW Written To Level	0
PW Wisdom ID	{576DD72C-A17A-4FB3-8A7F-00EB29013624}
<b>Education Dimension Table</b>	
Education Level All	NA
Education Level Education L	[Education Level].[All Education Level].[High School Degree]
PW Written To Level	00001
PW Wisdom ID	{576DD72C-A17A-4FB3-8A7F-00EB29013624}
<b>Gender Dimension Table</b>	
Gender All	NA
Gender Gender	[Gender].[All Gender].[M]
PW Written To Level	1
PW Wisdom ID	{576DD72C-A17A-4FB3-8A7F-00EB29013624}
<b>Measure Dimension Table</b>	
Measures MeasuresLevel	[Measures].[Sales Count]
PW Written To Level	0
PW Wisdom ID	{576DD72C-A17A-4FB3-8A7F-00EB29013624}

Entries from some Dimensional tables, recording the dimensional context of the Information Nugget

FIG. 9B



Combination of various contexts automatically and uniquely catalog the Information Nugget

FIG. 9C



16/23

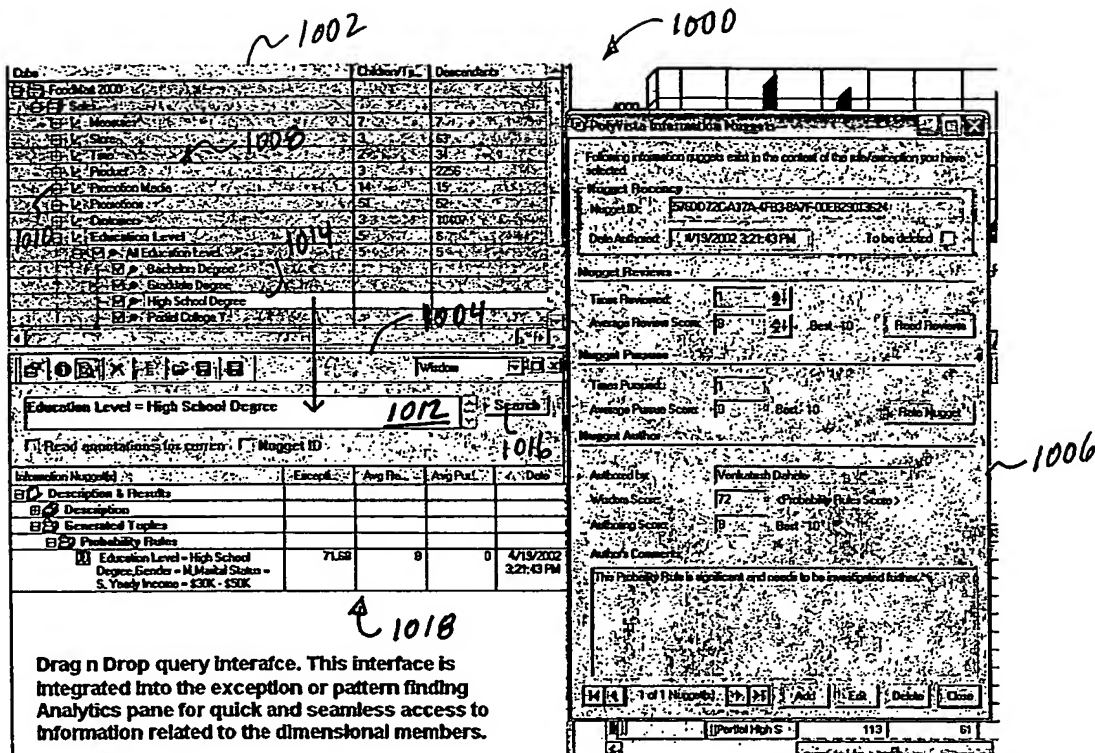


FIG. 10A

```

SELECT *
FROM
  [FoodMart_2000_Sales_Fact]          -- Table holding the Analytical Content of Information Nugget.
INNER JOIN
  [FoodMart_2000_Sales_Education_Level] -- Table holding the Education Level Dimensional context.
ON
  ([FoodMart_2000_Sales_Fact].[PW_Wisdom_ID]
  =
  [FoodMart_2000_Sales_Education_Level].[PW_Wisdom_ID])
INNER JOIN
  [FoodMart_2000_Sales_Time]          -- Table holding the Time Dimensional Context Specified in the
  -- the users Constraint Profile
ON
  ([FoodMart_2000_Sales_Fact].[PW_Wisdom_ID]
  =
  [FoodMart_2000_Sales_Time].[PW_Wisdom_ID])

WHERE
  -- User specified dimensional member
  (([Education_Level_Education_Level]='[Education Level].[All Education Level].[High School Degree]'))
AND
  -- Only Show Information Nuggets that are not marked for deletion.
  (([PW_Deletion_Flag]= 0))
AND
  -- Security Constraint for Venkatesh Debale transparently added.
  (([FoodMart_2000_Sales_Time].[Time_Year] <> '[Time].[1998]'))

ORDER BY
  -- Default Sort Order for Information Nuggets
  [FoodMart_2000_Sales_Fact].[PW_Average_Review_Score]DESC ,
  [FoodMart_2000_Sales_Fact].[PW_Average_Purpose_Score]DESC ,
  [FoodMart_2000_Sales_Fact].[PW_Date_Time]DESC
  
```

SQL Query Automatically Generated per the Dimensional Members Selected in the Search Box and users Security Constraints Profile.

FIG. 10B

17/23

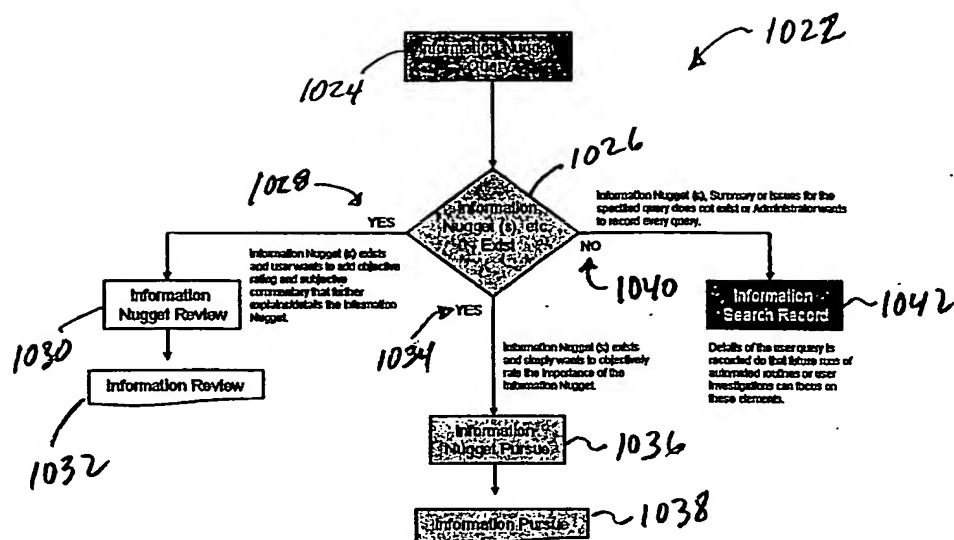


FIG. 10C

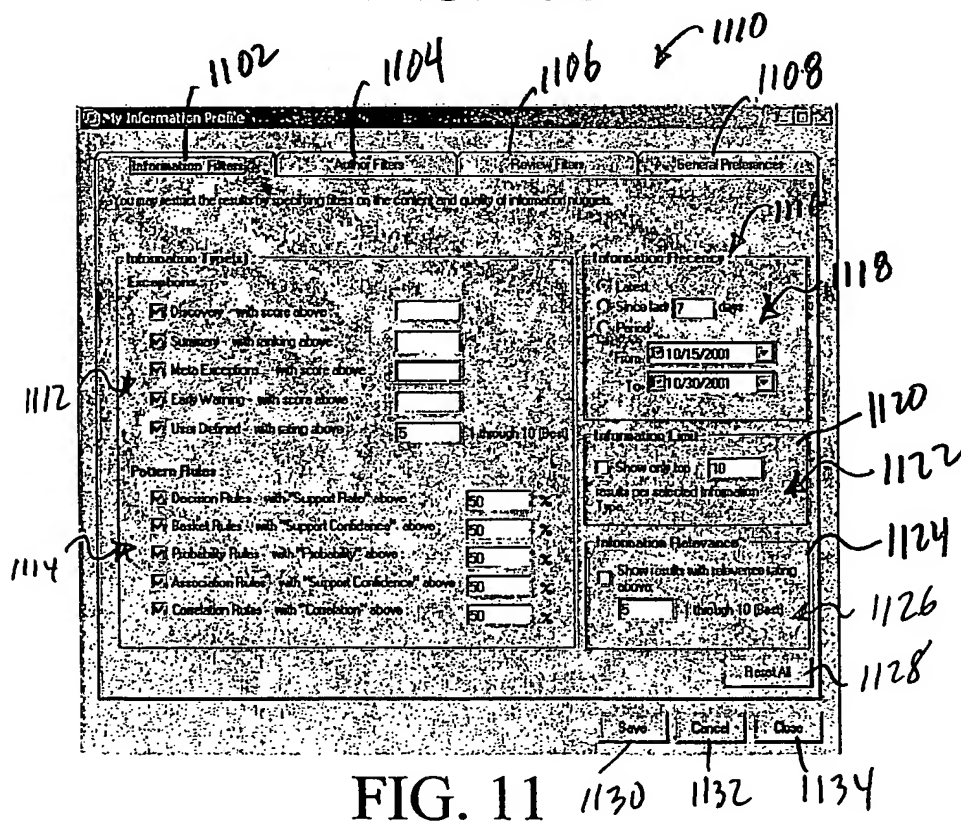


FIG. 11

18/23

- 1200

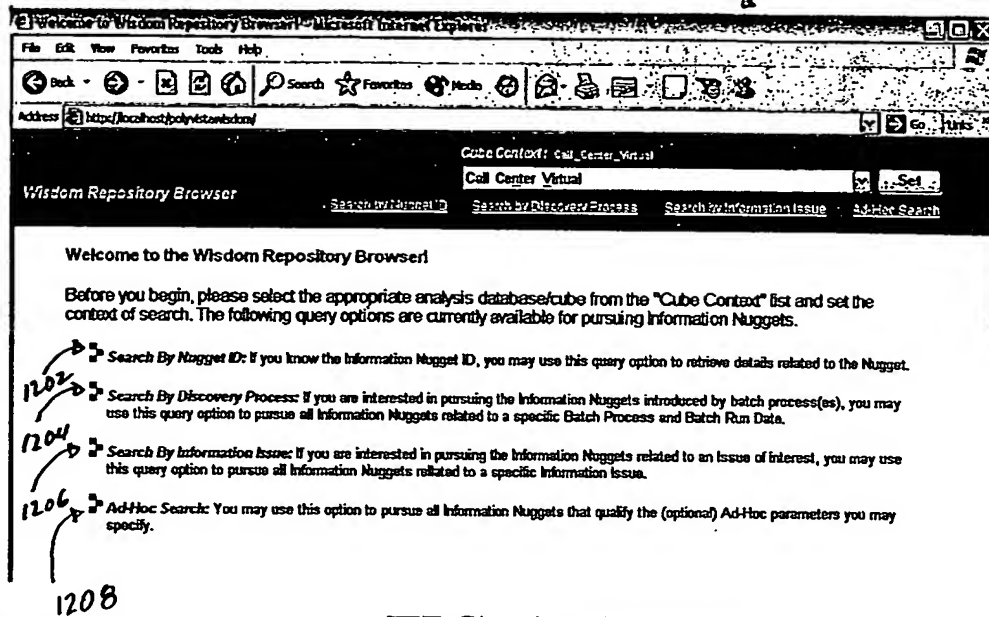


FIG. 12A

1210

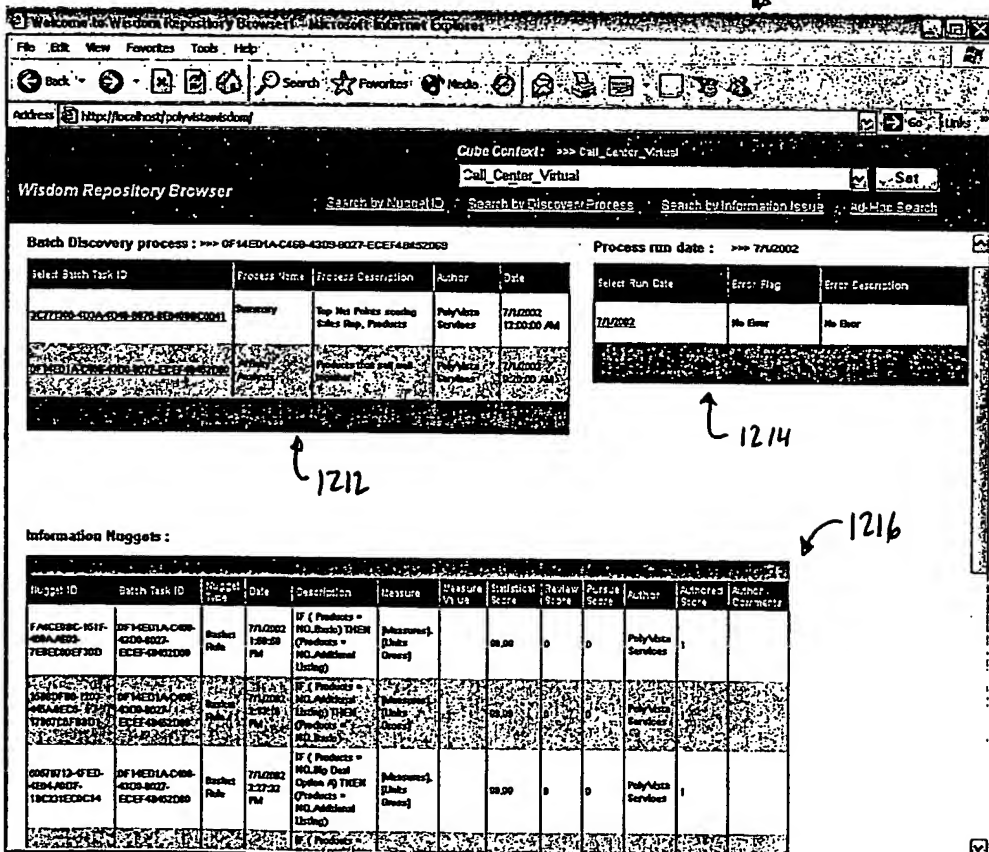


FIG. 12B

19/23

1218

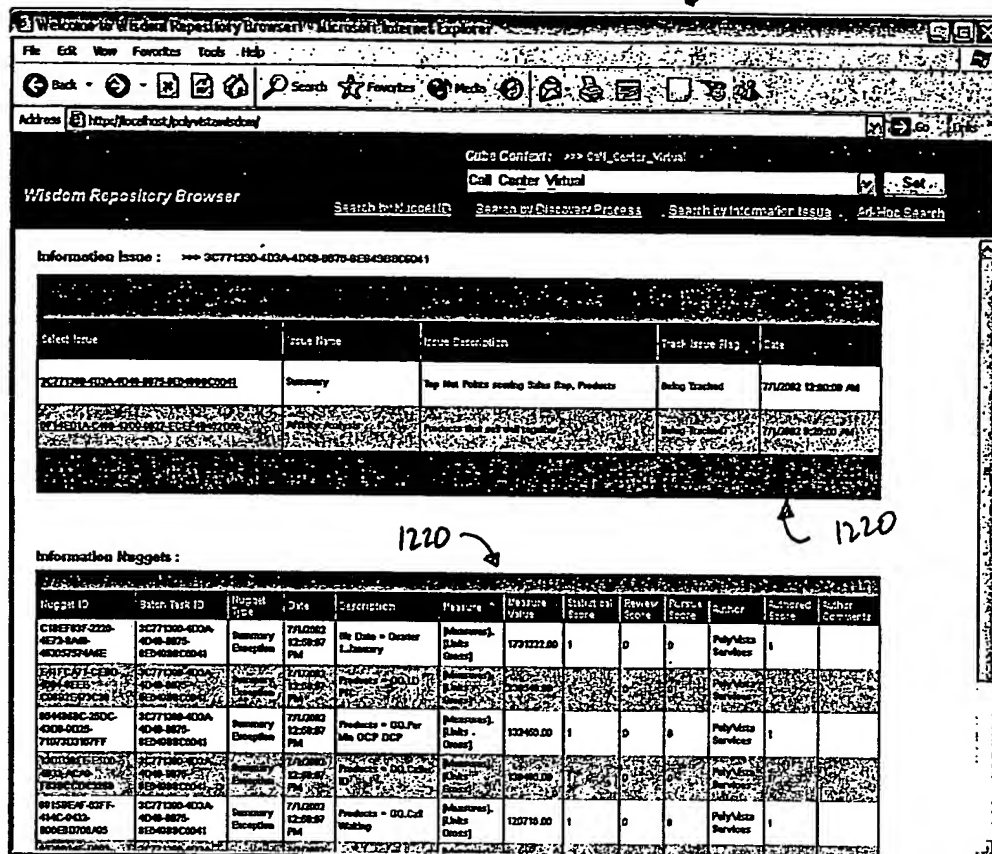





FIG. 12C

**20/23**

Address: <http://localhost:8080/virtualcenter>   

Cube Context: >>> [Virtual\\_Center\\_Virtual](#)

Wisdom Repository Browser [Search by context](#)

[Search by Facets](#) [Print page](#) [Search by information on screen](#) [Set Up Search](#)

**Specify Search Query Filters (Optional) -**

**Information Types (Optional):**

Exceptions -

☐ Discovery - with score between:  To

☐ Summary - with ranking above:

Pattern Rules -

☐ Neutral Rules - with "Support Confidence" above:

☐ Association Rules - with "Support Confidence" above:

☐ Probability Rules - with "Probability" above:

**Information Relevance (Optional):**

☐ Information nuggets with person score above:

☐ Information nuggets with review score above:

☐ Information nuggets with authored score above:

**Information Query Keywords (Optional):**

☒ Use Keywords

Example: "MCKENZIE"

**Information Query Options (Optional):**

Apply the following constraints simultaneously ("ANDing"):

☐ Information Relevance

☐ Information Rarity

[Get Nuggets](#)

Wisp ID	Wisp Task ID	Wisp Type	Wisp Data	Description	Wisp Name	Wisp Value	Wisp Unit	Wisp Score	Wisp Rank	Wisp Author	Wisp Score	Wisp Comments
2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Summary	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Info Rule - MCKENZIE, DAVID, DOONCE, JONES	(Name), (Data Group)	100.00	1	0	0	PolyData Service	1	
2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Summary	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Info Rule - MCKENZIE, DAVID, DOONCE, JONES	(Name), (Data Group)	100.00	1	0	0	PolyData Service	1	
2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Summary	2F8C8B-88F8-4D8B-88F8-88F8F8F8F8	Info Rule - MCKENZIE, DAVID, DOONCE, JONES	(Name), (Data Group)	100.00	1	0	0	PolyData Service	1	

FIG. 12D

Address: [http://0:13doVaskoldVexploratory.html]

# Food Mart Inc.

## Marketing info database

### Marketing department

- Home
- Register
- My Page
- ???
- ???
- Contacts

Select:  Measure:

12b2

Boolean | Pricedlines | DB Structure

What is the profit for drink in CA?

Show  results per page

FIG. 12E

21/23

Search results

What is the profit for drink in CA?

1.

1266 →

→ 1268

→ 1264

Database: FoodMart Cube: Warehouse Measure: Profit

☒ [Product].[Drink]

☒ [Store].[USA].[CA]

☒ [Warehouse].[USA].[CA]

☐ [Time].[1997], [Time].[1998]

☐ [Store Size in SQFT]

☐ [Store Type]

2.

Database: FoodMart Cube: Sales Measure: Profit

☒ [Product].[Drink]

☒ [Store].[USA].[CA]

☐ [Customers].[USA].[CA]

☐ [Time].[1997], [Time].[1998]

☐ [Promotion Media].[All Media]

☐ [Promotions].[All Promotions]

☐ [Education Level].[All Education Levels]

☐ [Gender].[All Genders]

☐ [Marital Status].[All Marital Status]

☐ [Store Size in SQFT].[All]

☐ [Store Type].[All]

☐ [Yearly Income].[All Yearly Income]

Database: FoodMart Cube: Warehouse and Sales Measure: Profit

☒ [Product].[Drink]

☒ [Store].[USA].[CA]

☐ [Warehouse].[USA].[CA]

☐ [Customers].[USA].[CA]

☐ [Time].[1997], [Time].[1998]

☐ [Store Size in SQFT]

☐ [Store Type]

☐ [Promotion Media]

☐ [Promotions]

☐ [Education Level]

☐ [Gender]

FIG. 12F

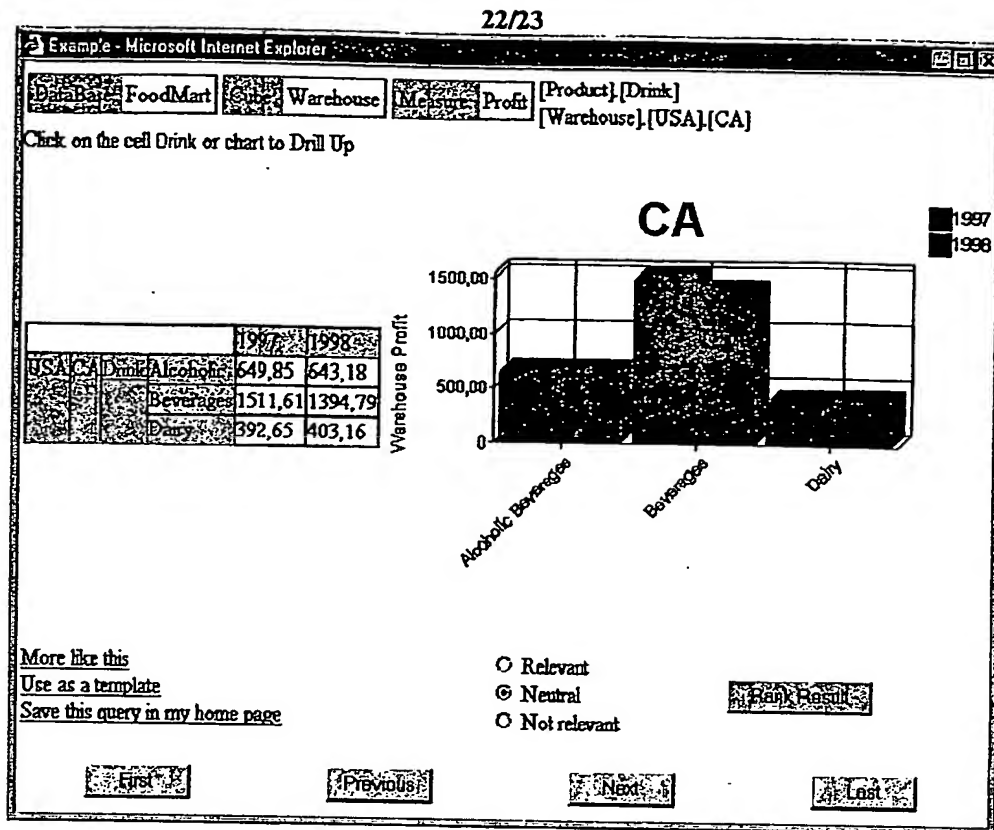


FIG. 12G

1270

1300

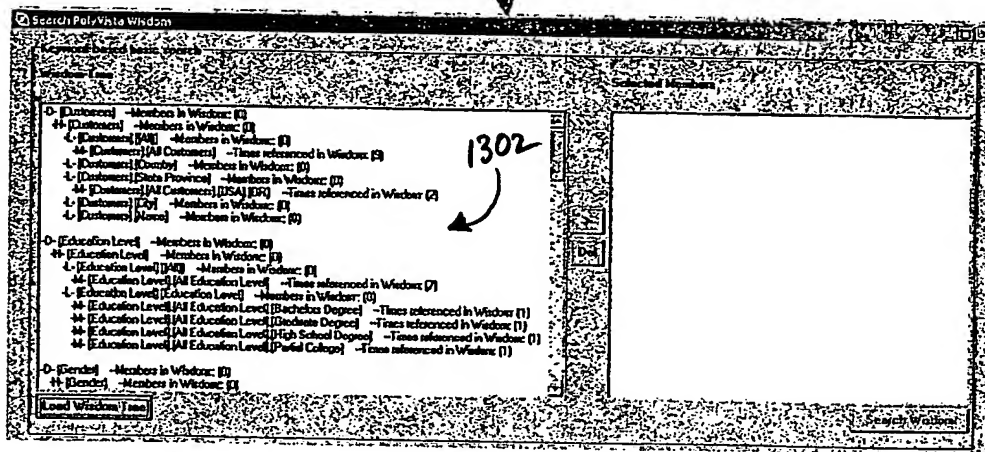


FIG. 13

**23/23**

[illegible]

### Data Mining (Basket Analysis) Task Definition

**Backlist Photos** 1942-1945: 1942-1945

### Description & Results

ID	Description	Status	Date	Time	Location	Remarks
1	...	...	...	...	...	...

- ☒ SERVER = Localhost
- ☒ DATABASE = FoodMart 2000
- ☒ CUBE = Sales
- ☒ MEASURE = {Measures}[Sales Count]
- ☒ TRANSACTIONS:  
( {Dimensions}[Name], {Time}[Month], {members} )
- ☒ FILTERS:  
( {Product}[Product Subcategory], {members} )
- ☒ MUST INCLUDE:
- ☒ FILTERS:
- ☒ MIN SUPPORT = 20.0
- ☒ MIN CONFIDENCE = 30.0
- ☒ MIN IMPROVEMENT = 1.0

### Repository Product Dimension Table

Here the items parameters are being modified to ignore some of the Product Subcategory members already contained in the Nugget Repository.

EXCEPT ((Product), (Product Subcategory), MEMBERS), (Product), (All Products), (Drink), (Alcoholic Beverages), (Beer and Wine), (Beer), (Product), (All Products), (Non-Consumable), (Health and Hygiene), (Bathroom Products), (Mouthwash))

**"Transactions", "Filters", "Must Include" and "Items" parameters can be customized based on the contents of the Nugget Repository.**

FIG. 14



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**